Trustworthy interactive learning, a story of fairness, robustness and safety

Romain Camilleri

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:

Kevin Jamieson, Chair
Maryam Fazel
Jamie Morgenstern
Lalit Jain

Program Authorized to Offer Degree:
Paul G. Allen School of Computer Science and Engineering

University of Washington

Abstract

Trustworthy interactive learning, a story of fairness, robustness and safety

Romain Camilleri

Chair of the Supervisory Committee:

Professor Kevin Jamieson

Department of Computer Science and Engineering

In an era characterized by the rapid evolution and widespread adoption of machine learning, the imperative to balance innovation with responsible deployment has never been more pressing. This thesis embarks on a multifaceted exploration of efficient and interactive machine learning deployment, with a focus on advancing methodologies for the development of trustworthy and reliable models. Beginning with an in-depth analysis of high-dimensional experimental design and kernel bandits, it investigates the nuances of modeling smooth reward functions in Reproducing Kernel Hilbert Spaces (RKHS) and introduces novel algorithms for regret minimization and pure exploration tasks. This exploration extends to level set estimation in kernel bandits, shedding light on nearly optimal algorithms and their performance characteristics. Building on these foundational principles, the thesis progresses to examine selective sampling for online best-arm identification, enabling to leverage unlabeled and labeled data optimally. Further, it delves into the challenge of non-stationary linear bandits, emphasizing robustness in dynamic environments. Additionally, the thesis explores integrating safety and fairness constraints into the active learning paradigm, empowering the alignment of model development with ethical considerations and real-world constraints. This effort enhances the reliability of machine learning models and contributes to the advancement of responsible AI. In conclusion, this thesis offers a comprehensive exploration of efficient interactive learning deployment, presenting novel insights, methodologies, algorithms, and perspectives to foster responsible innovation in the field.

Acknowledgements

I would like to express my heartfelt appreciation to my advisor, Professor Kevin Jamieson. I am deeply grateful to him for taking a chance on me early in my PhD journey. His love for bandits and AI in general profoundly influenced me a lot. His deep technical intuition, evident through insightful questions and innovative solutions, has significantly shaped my thinking. Beyond his expertise, Kevin has been a true mentor, helping me grow scientifically, professionally, and personally. His passion for research, endless ideas, and unwavering perseverance have been a tremendous source of motivation. Kevin's clarity of thought and dedication to exploring various fields continue to inspire me. His kindness, respectfulness, thorough guidance, and unwavering support have been invaluable throughout my journey.

Throughout my PhD, I have had the privilege of closely collaborating with Professors Maryam Fazel, Lalit Jain, and Jamie Morgenstern. Our regular interactions have encouraged me to adopt a more open-minded and holistic approach while maintaining attention to detail. I am deeply grateful to all of them, as well as to my GSR, Professor Alexandr Aravkin, for forming such a responsive, supportive, and guiding committee, and for providing comprehensive feedback. Additionally, I am grateful to the machine learning faculty — Professors Pang Wei Koh, Ludwig Schmidt, Simon Du, and Sewoong Oh — for fostering an exceptional research environment in the CSE department.

I would like to express my gratitude to my internships mentors: Dr. Zohar Karnin during my internship at AWS, and Dr. Moises Goldszmidt during my internship at Apple. Each internship was crucial to my development as a researcher — they taught me how to conduct real applied research, and provided inspiration for research questions drawn from applied problems.

My academic journey was greatly enriched by my academic siblings: Jennifer Brennan, Andrew Wagenmakers, Yifang Chen, Arnab Maiti, Artin Tajdini, Julian Katz-Samuels, Jifan Zhang, Yuhao Wan, and Stephen Mussmann. Their infectious enthusiasm and inspiring work ethic made my experience incredibly rewarding. I have learned immensely from each of them and could not have asked for better companions. Special thanks to Andrew for being an excellent collaborator and an even better friend.

I sincerely thank Elise Dorough, Joe Eckert and the graduate advising team for their support throughout graduate school.

I acknowledge my co-authors (in chronological order): Kevin Jamieson, Julian Katz-Samuels, Blake Mason, Lalit Jain, Robert Nowak, Subhojyoti Mukherjee, Zhihan Xiong, Maryam Fazel, Andrew Wagenmaker, and Jamie Morgenstern.

I have been fortunate to share this PhD journey with wonderful individuals who have become dear friends: Andrew Wagenmaker, Oscar Sprumont, Krishna Pilllutla, Amanda Baughan, Sally Dong, Swati Padmanabhan, Pascal Sturmfels, and the UW ML group, especially Claire Zhang, Zhaoqi Li, Tanner Fiez, Rahul Kidambi, Sam Ainsworth, Alexander Greaves-Tunnell, Rachel Hong, Corinne Jones, Ramya Vinayak, John Thickstun, Mitas Ray, Omid Sadeghi, Divyansh Pareek, Daniel Jiang, Gavin Brown, Jonathan Hayase, and Erfan Loghmani. I am also grateful to Jean-Pierre, Maria, Nathan, Siddarth, Ella, Emmett, Nikhil, Clem, Agathe, and Quentin for their friendship and good times throughout my Seattle journey.

I'm grateful to all my teachers and mentors who helped shape who I am today. I want to give special thanks to Professors Florian Schafer and Houman Owhadi for guiding my early research. The lessons they taught me remain invaluable.

I am deeply thankful to the friends I made before moving to Seattle and to my family who taught me life's joys: my parents, my siblings, and all my grandparents, cousins, aunts, and uncles for their unwavering support. Most of all, I am grateful to my incredible partner, Alex, and her dogs, Murphy and Snoopy, for their constant support. Their sacrifices have made my PhD journey possible, and I share this achievement with all of them.

Contents

Chapter 1

Introduction

Artificial intelligence is witnessing an extraordinary surge, marked by unprecedented advancements spanning natural language processing, computer vision, speech recognition, and pioneering fields like protein structure prediction and vaccine development (LeCun et al., 2015; Hannun et al., 2014; Jumper et al., 2021; Keshavarzi Arshadi et al., 2020). This resurgence owes its momentum to groundbreaking strides in large-scale models, notably deep neural networks and large language models. Leveraging the wealth of expansive datasets, state-of-the-art machine learning architectures now empower AI systems to achieve remarkable feats, from mastering strategic games like Go (Mnih et al., 2015; Silver et al., 2016) to facilitating robot-assisted surgeries with unparalleled precision (Kassahun et al., 2016).

In an era defined by the rapid advancement and pervasive adoption of machine learning, striking a delicate balance between innovation and responsible deployment has become an intellectual frontier that demands our utmost attention. One of the primary challenges in this landscape lies in ensuring the availability, quality and integrity of the data upon which these machine learning systems rely. As these powerful tools seamlessly integrate into our everyday lives, from automated decision-making in marketing (Schwartz et al., 2017) and healthcare (Esteva et al., 2019) to autonomous vehicles (Kiran et al., 2021) and personalized recommendations (Bouneffouf et al., 2012), the need for efficient and informative data collection, robust policies and models is more crucial than ever.

The scientific community has been urgently mobilized to combat the COVID-19 pandemic (Gurung et al., 2021; Liu et al., 2021; Piccialli et al., 2021). Understanding COVID-19's pathobiology could help identify potent antivirals by revealing new viral pathways. Computational methods, particularly machine learning-based models, have become invaluable for discovering candidate drugs and vaccines in silico. These models can predict inhibitor candidates based on structural data, aiding the search for effective treatments (Keshavarzi Arshadi et al., 2020; Mouchlis et al., 2021). As we will see next, advances in artificial intelligence highlight its potential in developing COVID-19 therapeutics.

In the subsequent sections of this chapter, we will address the task of predicting whether a patient will respond well to a therapeutic treatment (e.g., drugs, vaccines). Through this illustrative example of a machine learning problem, we will explore the emerging challenges of data collection and the reliability of contemporary machine learning systems. Additionally, we will outline how this thesis addresses these challenges. Finally, we will conclude this chapter by summarizing the contributions made across different chapters of this dissertation in confronting these pertinent issues.

1.1 Two Challenges

1.1.1 Data collections

The challenge of data collection is multifaceted, encompassing the costs of acquiring data and ensuring its quality and relevance. This section delves into the intricacies of effective data gathering, addressing the complexities inherent in modern data collection processes.

Data collection often incurs substantial expenses, particularly when involving real-world experiments or specialized expertise. For example, predicting whether a patient with SARS-CoV-2 will respond well to a therapeutic treatment requires collecting extensive examples from medical trials, observing the effects of administering the drug to patients. Tasks like medical image segmentation necessitate highly trained professionals, further increasing data collection costs (Gao et al., 2012; Guo et al., 2019). Additionally, as datasets grow in size, so do the associated expenses, as illustrated by monumental datasets like ImageNet, which have driven recent advancements in machine learning (Russakovsky et al., 2015).

The study of data collection techniques traces back to the dawn of modern statistics. Optimal experimental design, dating back at least a century, stands as a testament to the enduring quest for efficient data gathering methodologies (Smith, 1918).

Exploring data collection requires a deep understanding of its complexities and potentials. This thesis examines the increasing costs of data gathering and explores new techniques such as active learning within the fields of contemporary machine learning and statistics. By delving into past knowledge, addressing present challenges, and proposing pathways for improved data collection approaches, we aim to catalyze significant progress in the field. This involves not only understanding the challenges but also embracing opportunities for innovation and optimization.

1.1.2 Model reliability

As machine learning applications continue to evolve in complexity and scale, it becomes increasingly crucial to exercise foresight and prudence. Policies and models crafted without meticulous consideration can give rise to unforeseen vulnerabilities, such as the illusion of false confidence, unintended performance degradation, adoption of risky strategies, a propensity for false discoveries, and the perpetuation of social inequities. These missteps not only undermine the trust placed in machine learning systems but also diminish the potential benefits they can provide (House; Barocas and Selbst, 2016; Lum and Isaac, 2016; Buolamwini and Gebru, 2018; Angwin et al., 2022). This dissertation aims to address these critical issues, contributing to the development of more trustworthy machine learning.

Managing the illusion of false confidence is crucial, as model misspecification poses significant risks (Camilleri et al., 2021a; Mason et al., 2021). Machine learning models, despite their remarkable capabilities, are only as good as the data they are trained on and the assumptions they make. For example, consider the modeling of the effectiveness of a COVID-19 vaccine. If the model incorrectly assumes that the vaccine's effectiveness is uniform across ages, it could lead to biased and potentially overestimated effectiveness rates. This misspecification might occur if the model fails to account for varying immune responses among different age groups, preexisting conditions, or genetic backgrounds. Consequently, public health decisions based on such a flawed model could result in the misallocation of vaccine resources, underprotection of vulnerable populations, and ultimately, catastrophic failures in managing the pandemic. We will see in this work the importance of approaching model development with a healthy skepticism and a thorough understanding of the limitations and potential biases that may be inherent in the data.

One of the key challenges lies in managing the decay of performance, ensuring that policies and models

remain reliable and effective over time. A model that excels initially may stumble when confronted with evolving data patterns or unanticipated scenarios. To prevent this degradation, more flexible and adaptable strategies, robust and generalizable model designs, and continuous monitoring become paramount (Abbasi-Yadkori et al., 2018; Xiong et al., 2024). For example, consider again the modeling of the effectiveness of a COVID-19 vaccine. If initial models are built on early clinical trial data that do not fully capture the diversity of the population or the virus's potential to mutate, their predictions may become less accurate as new variants emerge or as different demographic groups receive the vaccine. This time variation could lead to overestimating the vaccine's effectiveness in the general population or underestimating the need for booster shots in specific groups. Such changes could result in misguided public health policies, where insufficient resources are allocated to high-risk communities, leading to higher infection rates and preventable deaths. Moreover, failing to anticipate a decline in vaccine effectiveness against new variants might delay necessary updates to vaccination strategies, prolonging the pandemic and increasing the strain on healthcare systems. Regular validation and updating of models with new data and insights are essential to maintain their reliability and accuracy in guiding critical health interventions. Continuous monitoring and adaptation allow for the detection of performance decay and enable timely adjustments (Abbasi-Yadkori et al., 2018; Xiong et al., 2024), ensuring that health policies remain effective in protecting public health. These approaches will be extensively covered in this work.

Managing the adoption of risky strategies in machine learning requires the incorporation of a risk function into the model development process. A risk function allows decision-makers to quantify and evaluate the potential consequences and trade-offs associated with different courses of action. By explicitly considering the risks and benefits, this thesis demonstrates how decision-makers can make informed choices regarding the deployment and utilization of machine learning models (Pacchiano et al., 2020; Wang et al., 2022; Camilleri et al., 2022). For example, consider the modeling of the effectiveness of a COVID-19 vaccine. In this context, a risk function could help quantify the potential side-effects of the vaccine. Omitting side-effects might lead to premature decisions, resulting in preventable deaths. By integrating a risk function, decision-makers can better assess the trade-offs between efficacy and risks, and implement strategies that minimize negative outcomes while maximizing public health benefits. Incorporating risk functions into model development ensures that potential adverse effects are systematically evaluated, enabling more robust and responsible deployment of machine learning models. This approach will be extensively covered in this work, highlighting its critical role in guiding effective and safe health interventions.

Additionally, controlling the False Discovery Rate (FDR) is essential in developing trustworthy machine learning models. High FDR can lead to false positive errors with significant consequences, such as waste of resources due to inflated scientific findings or misleading medical treatments. Therefore, employing methodologies to set appropriate decision thresholds, statistical adjustments, and validation strategies are crucial to minimize the occurrence of false discoveries and enhance the reliability of machine learning systems (Jain and Jamieson, 2020; Camilleri et al., 2022). For example, consider the modeling of the effectiveness of a COVID-19 vaccine. If the FDR is not adequately controlled, the model might falsely identify a vaccine as highly effective when it is not. This could lead to inappropriate decisions for vaccination. Such false positives in vaccine efficacy modeling could result in wasted resources, incorrect public health strategies, and ultimately, higher infection and mortality rates. By implementing advanced techniques to maintain a low FDR, decision-makers can ensure that vaccine efficacy models are more accurate and reliable. This, in turn, leads to better-informed public health policies, optimal resource allocation, and improved outcomes in managing the pandemic. This dissertation presents advanced techniques for efficiently maintaining low FDR in ML applications, highlighting their importance in ensuring the reliability and trustworthiness of machine learning systems.

Furthermore, addressing the ethical dimension of machine learning deployment and the potential perpetuation of social inequities is crucial. Machine learning models can inadvertently inherit biases from the data they are trained on, resulting in unjust outcomes and exacerbating existing disparities. Conscientious efforts, such as the ones presented in this dissertation, must be made to curate diverse and representative datasets, employ fairness metrics, and foster inclusivity throughout the model development process, ensuring that these powerful tools align with the principles of justice and equality (Barocas et al., 2017; Camilleri et al., 2023). For example, consider the modeling of the effectiveness of a COVID-19 vaccine. If the training data used for the model predominantly comes from a specific demographic group, the resulting model might not accurately reflect the vaccine's effectiveness across diverse populations. This can lead to biased health recommendations and unequal access to effective treatments. For instance, minority communities might be unfairly deemed less in need of vaccination based on flawed model outputs, perpetuating health disparities and increasing vulnerability to the virus. By curating diverse and representative datasets and applying fairness metrics, we can develop more equitable vaccine efficacy models that ensure all populations are fairly represented and protected. This approach not only improves the accuracy and fairness of the models but also aligns with broader ethical principles of justice and equality. This dissertation explores these conscientious efforts, demonstrating how they can help mitigate bias and promote inclusivity in machine learning applications.

In this pivotal moment of machine learning proliferation, we are presented with a unique opportunity to shape the future of this transformative technology. By recognizing the potential pitfalls and challenges that accompany its rapid advancement, and by actively engaging in the development of thoughtful models and policies, this dissertation steers the trajectory of innovation towards a future where machine learning empowers humanity while upholding our fundamental values. Prioritizing safety, robustness, and ethical frameworks becomes paramount in this journey. Let this work embrace the challenges ahead, guided by a shared commitment to responsible machine learning, ensuring a future that not only balances technological progress but also safeguards the well-being of individuals and fosters inclusivity within society as a whole.

1.2 Contributions

The pure exploration problem originated within the framework of multi-arm bandits (MAB) (Robbins, 1952; Even-Dar et al., 2002). In this problem, a learner sequentially selects an arm $i \in [K]$ and receives a reward from an unknown distribution specific to that arm of mean $[\theta_*]_i$. The learner's objective is to efficiently allocate resources to identify the arm with the highest expected reward, represented as $\arg\max_{i\in[K]} \mathbb{E}[[\theta_*]_i]$. For example, consider the modeling of the effectiveness of different COVID-19 vaccines as a multi-armed bandit problem. Each arm $i \in [K]$ represents a different vaccine candidate, and the reward $[\theta_*]_i$ corresponds to the vaccine's efficacy based on clinical trial data. The learner's task is to determine which vaccine has the highest expected effectiveness. Efficiently identifying the most effective vaccine is crucial for optimal resource allocation, such as prioritizing manufacturing and distribution efforts. Misallocating resources to a less effective vaccine could lead to higher infection rates and prolonged pandemic conditions. By applying the principles of this sequential decision-making framework, we can systematically and effectively determine the best vaccine candidate, ultimately guiding public health decisions and maximizing societal benefits.

More recently, the stochastic linear bandit problem (LB), a variant of the MAB setup, has garnered significant attention since its introduction by (Auer et al., 2002a). In the stochastic LB setting, the input space X is a subset of $\mathbb{R}^d$. When the learner selects an arm x, they receive a noisy measurement y of the reward function $f(x)$ that is a linear combination of x and an unknown parameter $\theta_* \in \mathbb{R}^d$, specifically $y = f(x) +$

$\varepsilon = x^\top \theta_* + \varepsilon$ where ε is some centered gaussian iid noise. The linear structure of the problem implies that pulling an arm provides information about the parameter θ_* and, indirectly, about the values of other arms. Consequently, rather than estimating the mean rewards for K arms, the task shifts to estimating the d features of θ_*. For example, consider the modeling of the effectiveness of different COVID-19 vaccines using a stochastic linear bandit approach. Each arm $x \in X$ represents a different combination of demographic and clinical features, such as age, preexisting conditions, and genetic backgrounds. The reward y corresponds to the observed vaccine efficacy for those specific features, incorporating noise from clinical trial variability. The unknown parameter θ_* represents the true underlying factors influencing vaccine efficacy. By selecting various arms (i.e., different demographic and clinical combinations), the learner gathers information about θ_*, improving the overall understanding of how different factors affect vaccine performance. This approach allows for more efficient identification of which demographic groups benefit most from certain vaccines, guiding targeted vaccination strategies and ensuring that public health resources are allocated effectively and equitably.

Thus, MAB and LB strategies differ in their approach. In MAB, arms are discarded once their sub-optimality becomes evident, while in LB, even sub-optimal arms provide valuable information about the parameter vector θ_*, thereby improving estimation accuracy. LB focuses on refining estimates along dimensions that differentiate remaining arms. The works covered in this dissertation build on this framework and explore the following directions: kernel bandits, robustness in bandits via model misspecification, best of both worlds, safety and transductive bandits-based active learning for accuracy, and accuracy with false discovery rate (FDR) or fairness constraints.

1.2.1 Kernel bandits

To address large or continuous domains, modern bandit approaches model and exploit the problem structure, often manifested as correlations in rewards of "similar" actions. This is particularly relevant in practical scenarios like COVID-19 vaccine development, where understanding correlations between different demographic and clinical groups is crucial. The key idea of kernelized bandits is to consider only smooth reward functions of a low norm belonging to a chosen Reproducing Kernel Hilbert Space (RKHS) of functions. Specifically, let the reward function f be modeled in an RKHS $\boldsymbol{H}$. Let $\phi : \mathbb{R}^d \mapsto \boldsymbol{H}$ be the feature map associated with the RKHS, such that $f(x) = \langle \theta_*, \phi(x) \rangle_{\mathcal{H}}$ for $x \in \mathcal{X}$.

Pure exploration and regret minimization. Although previous studies have addressed regret minimization for kernel bandits, our work (Camilleri et al., 2021a) introduced the pure exploration problem specifically for this context. For instance, in the context of COVID-19 vaccine development, pure exploration involves efficiently identifying the most effective vaccine by exploring different combinations of demographic and clinical features. Previous works by (Srinivas et al., 2009) and (Valko et al., 2013) proposed kernelized versions of the UCB algorithm (Auer et al., 2002a). In comparison to these works, we have designed a regret minimization algorithm that demonstrates comparable worst-case performance in terms of an *information gain* metric. Additionally, we have extended phased-elimination type algorithms (Fiez et al., 2019; Lattimore et al., 2020) to the kernel space to establish instance-dependent guarantees for both the regret minimization and pure exploration algorithms. For example, in vaccine development, this could mean designing algorithms that ensure precise and effective allocation of trials across different groups to maximize information gain. Notably, (Fiez et al., 2019) utilizes an optimal experimental design approach, which provides continuous allocation that needs to be rounded for selecting measurements to query. However, this rounding introduces an additive term proportional to the dimension of the space, which can be potentially infinite in our case. Therefore, extending instance-dependent optimal algorithms (Fiez et al., 2019) necessitated the development of new tools for performing experimental design in Kernel space, also

known as Bayesian experimental design (Derezinski et al., 2020; Alaoui and Mahoney, 2014). We establish a performance guarantee for our proposed Kernel experimental design by relating it to the notion of *effective dimension* introduced in (Derezinski et al., 2020; Alaoui and Mahoney, 2014).

Level set estimation. Our work (Mason et al., 2021) tackle the level set estimation problem for kernel bandits. In the context of COVID-19 vaccine development, level set estimation involves identifying demographic and clinical groups for which the vaccine efficacy exceeds a certain threshold, ensuring optimal and equitable distribution. Given a threshold value $\alpha \in \mathbb{R}$, the goal is to identify the set of points of the measurement space for which the value is above α, that is $G_\alpha := \{x \in \mathcal{X} : f(x) > \alpha\}$. Level set estimation is traditionally tackled by Bayesian optimization methods that typically use an acquisition function to minimize learner uncertainty. (Bryan et al., 2005) introduced Gaussian processes and the Straddle heuristic. (Gotovos, 2013) developed the LSE and LSE-imp algorithms with theoretical guarantees on sample complexities. (Bogunovic et al., 2016) connected Bayesian optimization with level set estimation, considering heteroscedastic noise. (Shekhar and Javidi, 2019) proposed an algorithm for the continuous domain, offering improved computational complexity and tighter sample complexity bounds. (Zanette et al., 2018) treated level-set estimation as a classification problem, introducing a new acquisition function. (Iwazaki et al., 2019) extended this work to enhance model robustness in quality control applications. These heuristics are known to work well in practice but often come with ad hoc guarantees and, at best, worst-case (minimax) performance. In contrast to the more heuristic approaches of Bayesian optimization methods, our work provides the first instance-dependent, non-asymptotic upper bounds on the sample complexity of level-set estimation. These bounds match the information-theoretic lower bounds established by (Kaufmann et al., 2016). To achieve these strong performances, we develop phased elimination algorithms with a general structure related to the instance-optimal algorithm for pure exploration tasks (Fiez et al., 2019; Camilleri et al., 2021a) and tailor the experimental design and elimination rule specifically for level set estimation. This ensures that we can for example efficiently identify groups for targeted vaccine distribution, thereby optimizing public health outcomes.

1.2.2 Bandits assumptions

Model misspecification. In certain real-life scenarios, such as the development of COVID-19 vaccines, the mapping of feature representations can be more complex than a simple linear process. For instance, accurately predicting vaccine efficacy might involve nonlinear interactions among demographic and clinical features. In such cases, a linear feature representation can provide an approximation of the value or reward functions, with a small and consistent error known as misspecification. Model misspecification is a commonly observed phenomenon and has been extensively studied in the context of regret minimization (Ghosh et al., 2017; Lattimore et al., 2020; Takemura et al., 2021; Dong and Yang, 2023; Camilleri et al., 2021a) and pure exploration (Camilleri et al., 2021a; Zhu et al., 2021; Mason et al., 2021; Réda et al., 2021), under the framework of *misspecified linear bandits*.

(Du et al., 2020) have demonstrated that searching for an action that is $\mathcal{O}(\varepsilon)$-optimal in these scenarios requires a minimum of $\Omega(\exp(d))$ queries. Nevertheless, if we lower our objective to finding an action that is not perfectly optimal but still reasonably close to it, there is still hope. (Lattimore et al., 2020) have discovered that it is possible to find a suboptimal action with an error of at most $\mathcal{O}(\varepsilon\sqrt{d})$ within $\text{poly}(d/\varepsilon)$ queries, where d represents the dimension of the feature vectors. In our work (Camilleri et al., 2021a), we tackle the feature space of potentially infinite dimension as *misspecified kernel bandits*, and demonstrate that the consequences of misspecification are comparable to the bias resulting from the necessary regularizer terms in the covariate estimates.

For example, when developing COVID-19 vaccines, even if the model used to predict vaccine efficacy is

not perfectly accurate (i.e., it is misspecified), we can still identify reasonably effective vaccines by accounting for the model's misspecification. Identifying the best vaccine might not be feasible if the difference in efficacy between two vaccines is smaller than the misspecification term. Our work (Camilleri et al., 2021a) is the first to characterize the difficulty of ε-good arm identification, directly extending the pure-exploration task to accommodate misspecified models. This means that despite the inherent complexity and potential inaccuracies in modeling vaccine efficacy, it is still possible to make informed and effective decisions about which vaccines to prioritize for further development and distribution.

Robustness to non-stationary measures. In the context of COVID-19 vaccine development, it is crucial to account for unknown variations in the reward distribution due to distribution shifts. For example, vaccine efficacy might vary over time due to emerging variants or changes in population immunity. Bandit applications that encounter such shifts require robust models to handle these dynamic conditions. Existing literature primarily concentrates on minimizing dynamic regret, which measures the deviation between the obtained reward and the reward of the best arm in each round t. (Garivier and Moulines, 2011) demonstrate that methods like (Auer et al., 2002b) can achieve a dynamic regret of approximately $\widetilde{O}(\sqrt{LT})$ when the number of distribution shifts, denoted by L, is known. A significant advancement is made by (Auer et al., 2019), who introduce an adaptive approach with the same dynamic regret, but without requiring knowledge of L. Recent contributions by (Chen et al., 2019) and (Wei and Luo, 2021) establish similar results in contextual bandit settings.

Other measures of non-stationarity, beyond L, have also been considered. For example, in the context of vaccine efficacy, non-stationarity might be quantified by total variation due to genetic drift of the virus, as discussed by (Chen et al., 2019). Additionally, (Suk and Kpotufe, 2021) propose the concept of severe shifts, which could correspond to significant changes in virus strains. While the aforementioned research focuses on constructing precise models of non-stationarity and developing tailored regret minimization algorithms, the "best of both worlds" (BOBW) approach that we cover next remains independent of such models.

The "best of both worlds" (BOBW) problem aims to design a bandit algorithm that can achieve optimal performance in both stationary and non-stationary scenarios, even without prior knowledge of the environment. This approach is particularly relevant for vaccine development, where conditions can change unpredictably. While most BOBW approaches focus on regret minimization (Bubeck and Slivkins, 2012; Seldin and Slivkins, 2014; Seldin and Lugosi, 2017; Auer and Chiang, 2016; Lee et al., 2021), (Abbasi-Yadkori et al., 2018; Xiong et al., 2024) specifically concentrate on BOBW for best-arm identification.

In (Xiong et al., 2024), we build on top of (Abbasi-Yadkori et al., 2018) by tackling the linear bandits case, while (Abbasi-Yadkori et al., 2018) only focused on the MAB case. Our algorithm is a fixed budget pure exploration algorithm—similar to the `sequential halving` (Karnin et al., 2013) and `Peace` algorithm (Katz-Samuels et al., 2020)—mixed with a G-optimal design strategy known to perform well in non-stationary settings (Lattimore and Szepesvári, 2020). This strategy is particularly useful in vaccine development for efficiently identifying the most effective vaccine under changing conditions. To establish the error rate bound, we use a set of virtual events based on the estimated gaps, related to (Abbasi-Yadkori et al., 2018), and an analysis of the subroutine RAGE-Elimination, similar to the ones by (Fiez et al., 2019; Katz-Samuels et al., 2020). This ensures that our approach remains robust even as the underlying conditions affecting vaccine efficacy evolve.

Safety constraints. In the context of COVID-19 vaccine development, it is crucial to consider safety constraints when identifying the most effective vaccine. Ensuring that a vaccine is not only effective but also safe for the population is paramount. Despite its importance, only a few existing studies have addressed the issue of identifying the best arm while considering safety constraints (Sui et al., 2015; 2020; Wang et al., 2022; Lindner et al., 2022; 2023; Camilleri et al., 2022).

The works by (Sui et al., 2015; 2020) tackle a general constrained optimization scenario, where the learner aims to minimize a function $f(x)$ over a domain $x \in D$, while having access to noisy samples of $f(x)$, represented as $f(x_t)+w_t$, and ensuring that a safety constraint $g(x_t) \leq h$ is satisfied for every queried point x_t. Although they provide an upper bound on the sample complexity, they do not provide a lower bound, and it has been demonstrated in (Wang et al., 2022) that their approach can be highly suboptimal.

On the other hand, (Wang et al., 2022) investigates the best-arm identification problem in multi-armed bandits. In their setup, at each time step t, they query a value $a_t \in A$ for a specific coordinate i_t and aim to identify the coordinate i^* that satisfies $a^*_{i^*}\theta_{i^*} \geq \max_i a^*_i\theta_i$, where $a^*_{i^*}$ denotes the largest value respecting the safety constraint: $a^*_{i^*} = \arg\max_{a\in A} a_i\theta_i$ s.t. $a\mu_i \leq \gamma$. Similar to (Sui et al., 2015; 2020), they require that the safety constraint $a_t\mu_{i_t} \leq \gamma$ holds during the learning process. While they establish matching upper and lower bounds and consider a slightly more general setting that allows for nonlinear (yet monotonic) response functions, they treat each coordinate as independent and do not permit information-sharing between coordinates, which is a key aspect targeted in the linear bandit setting.

In contrast, in (Camilleri et al., 2022), we allow the learner to query unsafe points during exploration and only require that a safe decision is made at the end of the process. This approach is particularly relevant in the development of COVID-19 vaccines. For example, during early stages of vaccine trials, some experimental conditions might initially seem unsafe but-in vitro-could lead to critical insights that ensure the final vaccine is both safe and highly effective. By allowing for such exploration, our methodology ensures a more thorough understanding of the vaccine's efficacy and safety, ultimately leading to better-informed public health decisions.

This approach ensures that the learner can explore a wider range of conditions, gathering valuable information that might otherwise be missed if constrained to safe queries only. By the end of the exploration process, the final decision on the best vaccine candidate is made with a comprehensive understanding of both its efficacy and safety, aligning with the stringent safety standards required in vaccine development. This balance between exploration and safety is crucial for the rapid yet responsible development of vaccines, especially in the face of a global health crisis like the COVID-19 pandemic.

Streaming setting. In the context of COVID-19 vaccine development, ensuring rapid and efficient identification of the best vaccine candidate is crucial. This scenario can be modeled as a streaming setting in linear bandits, where the decision-maker must balance between gathering enough evidence and minimizing the time and resources spent on trials. In our work (Camilleri et al., 2021b), we explore this challenge by adopting a novel online perspective for linear bandits best arm identification: instead of allowing the agent to select the next measurement they sample, we restrict their choice to whether they query the label of a given measurement, sampled independently and identically distributed (iid) at random from a given distribution.

The key difficulty of this selective sampling problem is to carefully trade off between the value of obtaining a label at the current time and waiting for a potentially more informative point to arrive in the stream. For example, in vaccine trials, this could mean deciding whether to analyze the efficacy data of a participant with common characteristics now or waiting for data from a participant with unique characteristics that could provide more insight. The agent faces a critical trade-off between the number of labeled samples they obtain and the point at which they gather enough evidence to make the best decision and stop collecting more samples. Efficiently managing this trade-off ensures that resources are optimally used, and crucial public health decisions are made promptly.

In (Camilleri et al., 2021b), we provide key insights into this trade-off between labeled samples and the duration of the sampling process. We establish an information-theoretic lower bound and propose an algorithm that achieves nearly optimal results. This balance is vital in the context of COVID-19 vaccine development, where delays in identifying the most effective vaccine can lead to prolonged health risks and

economic impacts. Our approach ensures that sufficient evidence is collected to make a well-informed decision while minimizing the time and resources expended.

We additionally provide experiments corroborating our findings, demonstrating the practical applicability of our algorithm in real-world scenarios. These experiments show that our approach can significantly reduce the number of samples needed while still accurately identifying the best vaccine candidate. This efficiency is particularly important in a pandemic situation, where time is of the essence and resources are limited. By applying our streaming setting methodology, we can expedite the vaccine development process, ensuring that effective vaccines are quickly identified and deployed to maximize public health benefits.

1.2.3 Fairness

Fairness In the context of COVID-19 vaccine development, ensuring that the vaccine distribution is fair across different demographic groups is crucial. Algorithmic fairness, which has gained significant interest in recent years, plays a vital role in this process, as evidenced by recent surveys (Barocas et al., 2017; Hort et al., 2022). Approaches to mitigate fairness disparities can be categorized into three lines of work: pre-processing, in-processing, and post-processing. Pre-processing aims to remove disparate impact by modifying the training data (Kamiran and Calders, 2012), while post-processing modifies already learned classifiers to improve fairness (Hardt et al., 2016).

In our work, we focus on in-processing techniques for bias mitigation, which involve modifying the learning process to build fair classifiers (Woodworth et al., 2017; Zafar et al., 2017b;a; Donini et al., 2020; Kallus and Zhou, 2019; Pleiss et al., 2017; Berk et al., 2017; Joseph et al., 2016; 2018; Agarwal et al., 2018; Cotter et al., 2018). Specifically, we are interested in approaches that treat fairness mitigations in classification as a constrained optimization problem (Agarwal et al., 2018; Donini et al., 2018; Camilleri et al., 2023). This focus is particularly relevant when determining the best vaccine candidate for various demographic groups, ensuring that no group is disproportionately disadvantaged by the decision-making process.

Fairness Violation Estimation A recent line of work has focused on developing the statistical foundations of in-processing for bias mitigation, which we will discuss next. Several studies (Ji et al., 2020; Miller et al., 2021; Barrainkua et al., 2023; Lum et al., 2022) have highlighted and proposed methods to address the issue of large variance in fairness violation estimates. For instance, (Ji et al., 2020) suggests a Bayesian method to obtain fairness measurements with reduced variance, and (Barrainkua et al., 2023) proposes a Bayesian inference strategy for more stable fairness measurements estimates. Similarly, (Miller et al., 2021) presents a multi-level modeling approach to construct tighter empirical confidence intervals for fairness measurements.

Notably, (Lum et al., 2022) identifies a potential statistical bias in fairness metrics commonly used in the literature. Instead of directly addressing this bias, they suggest alternative fairness definitions that have statistically unbiased estimators. In contrast, our work aims to tackle the bias issues directly for the fairness metrics typically of interest. Building on these prior works, our study (Camilleri et al., 2023) establishes concentration bounds on empirical fairness estimates in a rigorous manner and evaluates the performance of fairness estimates through experimental analysis.

In the application to COVID-19 vaccine development, these fairness considerations are crucial. For example, it is essential to ensure that vaccine efficacy is measured and reported fairly across different demographic groups, such as age, race, and socioeconomic status. By addressing fairness violations directly and providing reliable fairness estimates, our methodology helps ensure that vaccine distribution policies are equitable and just, preventing any demographic group from being unfairly prioritized or neglected. This

not only improves public trust in the vaccination process but also enhances overall public health outcomes by ensuring a fair and effective distribution of the vaccine.

1.2.4 Active learning

In the context of COVID-19 vaccine development, active learning has proven to be an invaluable tool. The high cost and logistical challenges associated with labeling vast amounts of clinical trial data make active learning essential for efficiently identifying the most promising vaccine candidates. Active learning allows for the strategic selection of the most informative data points to label, thus accelerating the development process by producing precise hypotheses with fewer labeled samples (Settles, 2011).

Active learning has been extensively studied over the past five decades, as evidenced by various research and surveys (Hanneke et al., 2014). Most active learning approaches select samples to label based on uncertainty measures such as entropy of predictions, margin, and disagreement (Cohn et al., 1994; Beygelzimer et al., 2009; Dasgupta, 2005; Balcan et al., 2006; 2007; Wang and Singh, 2015). These approaches have also been subject to analysis (Castro and Nowak, 2007; Dasgupta et al., 2009; Hanneke and Yang, 2014; Mussmann and Dasgupta, 2022). In the case of vaccine trials, selecting data points with the highest uncertainty helps identify the most informative patient responses, thereby optimizing the trial design and speeding up the path to an effective vaccine.

Recent breakthroughs have connected best-arm identification for linear bandits with classification, opening up new possibilities for active learning through experiment design (Katz-Samuels et al., 2021; Camilleri et al., 2021b; 2022; 2023). These approaches are based on experimental design, aiming to maximize information gain by querying points in the disagreement region. In the active learning algorithms described in (Camilleri et al., 2022; 2023), the agent computes a design by utilizing classifiers that have been trained with perturbed labels. This strategy is motivated by the randomized exploration approach introduced by (Kveton et al., 2019), where they demonstrate that randomized exploration functions as a form of posterior sampling.

Active learning with FDR constraint. The investigation of precision constraints in the adaptive context has received limited attention thus far. In COVID-19 vaccine trials, ensuring the precision of selected data points under constraints such as False Discovery Rate (FDR) is crucial for reliable results. Prior to our research (Camilleri et al., 2022), only a few studies, such as (Bennett et al., 2017; Jain and Jamieson, 2020), have explored this area. Our work has contributed to this field by developing methods that ensure precision in the adaptive sampling process, which is essential for making informed and accurate decisions during vaccine development.

Fair active learning. Ensuring fairness in the distribution and efficacy of COVID-19 vaccines is another critical challenge. Recent efforts have been devoted to achieving a favorable "fairness-error" trade-off in classifiers, given a label budget. Some notable works in this area include (Anahideh et al., 2021; Sharaf et al., 2022; Fajri et al., 2022). However, these works suffer from various limitations, such as poor generalization in handling fairness violations, minimal accuracy improvements compared to baseline methods, or limited capability to address standard group fairness metrics.

In (Camilleri et al., 2023), we propose a solution to the novel problem of reliably achieving fair active classification. It is important to note that our objective is to produce a classifier with fairness violations below a desired tolerance, as we recognize the significance of ensuring fairness in critical situations. In contrast, the aforementioned works primarily focus on quantifying the trade-off between fairness and accuracy, without guaranteeing that the resulting classifier falls within any specific tolerance level.

Other related studies have addressed similar problems under different fairness constraints. For example, (Shen et al., 2022; Cao and Lan, 2022a) focus on discovering classifiers that satisfy metric-fair constraints, while (Abernethy et al., 2021; Shekhar et al., 2021; Cai et al., 2022; Branchaud-Charron et al., 2021) target data collection methods for achieving min-max fairness. These metrics differ significantly from the group fairness metrics we consider, necessitating distinct methodologies.

In conclusion, just as strategic sample selection was crucial in the rapid and fair development of the COVID-19 vaccine, active learning principles and fairness considerations are critical in developing robust and equitable machine learning models. By ensuring that our algorithms can operate efficiently and fairly under various constraints, we can better handle real-world challenges like those encountered during the pandemic.

1.3 Thesis Outline

This thesis is structured to delve into various dimensions of interactive learning deployment, focusing on addressing critical challenges and advancing methodologies for responsible and effective model development. The following chapters outline the key contributions and findings of this research endeavor:

Chapter 1: Introduction
The introductory chapter sets the stage by delineating the landscape of machine learning deployment in contemporary times, emphasizing the importance of balancing innovation with responsible practices. It provides a comprehensive overview of the challenges and opportunities inherent in the rapid advancement of machine learning technologies.

Chapter 2: High-Dimensional Experimental Design and Kernel Bandits
This chapter explores the realm of high-dimensional experimental design and its application in kernel bandit settings. It delves into the intricacies of modeling smooth reward functions in Reproducing Kernel Hilbert Spaces (RKHS) and presents novel algorithms for regret minimization and pure exploration tasks.

Chapter 3: Nearly Optimal Algorithms for Level Set Estimation
Here, the focus shifts to the problem of level set estimation in kernel bandits, where the objective is to identify sets of points with rewards exceeding a certain threshold. The chapter introduces nearly optimal algorithms and analyzes their performance in terms of sample complexity and computational efficiency.

Chapter 4: Selective Sampling for Online Best-arm Identification
This chapter explores a novel perspective on online best-arm identification by introducing the concept of selective sampling. It investigates the trade-offs between labeled samples and sampling duration, presenting algorithms that achieve nearly optimal results in this context.

Chapter 5: A/B Testing and Best-arm Identification for Linear Bandits with Robustness to Non-stationarity
Building upon the foundation laid in earlier chapters, this section focuses on A/B testing and best-arm identification for linear bandits, specifically addressing robustness to non-stationarity. It examines algorithms that adaptively navigate dynamic environments while maintaining optimal performance.

Chapter 6: Active Learning with Safety Constraints

This chapter bridges the gap between active learning methodologies and best-arm identification, leveraging insights from the linear bandits problems addressed in previous chapters. It explores algorithms designed to facilitate safe decision-making while effectively utilizing unlabeled data. By integrating safety constraints into the active learning paradigm, this research endeavors to enhance the reliability and robustness of machine learning models in real-world applications.

Chapter 7: Fair Active Learning in Low-Data Regimes
The final chapter addresses the critical issue of fairness in machine learning, particularly in low-data regimes. It explores methodologies for achieving fair classification outcomes while operating within constrained label budgets, offering insights into the intersection of fairness and efficiency in active learning settings.

Chapter 8: Conclusion, Impact and Future Directions
The concluding section summarizes the key findings and contributions of the thesis, reflecting on the implications for the field of machine learning.

Chapter 2

High-Dimensional Experimental Design and Kernel Bandits

2.1 Introduction

This chapter studies a non-parametric multi-armed bandit game through the lens of experimental design. Fix a finite set of measurements $\mathcal{X} \subset \mathbb{R}^d$ and a function $\mu : \mathcal{X} \to \mathbb{R}$. We consider the following game between a learner and nature: at each time $t = 1 \dots T$, the learner requests $x_t \in \mathcal{X}$ and nature immediately reveals

$$y_t = \mu_{x_t} + \xi_t$$

where $\{\xi_t\}_{t=1}^T$ is a sequence of independent, mean-zero random variables with bounded variance. We are interested in two objectives:

Regret minimization In this setting, we evaluate the performance of an algorithm choosing actions $\{x_t\}_{t=1}^T$ by its cumulative regret: $R_T = \max_{x \in \mathcal{X}} \sum_{t=1}^T (\mu_x - \mu_{x_t})$.

Pure exploration in the PAC setting For a tolerance $\epsilon \geq 0$ and confidence level $\delta \in (0, 1)$, the aim of the learner in pure exploration is to sequentially take samples until a learner-defined stopping criterion is met, at which time the learner outputs an arm $\widehat{x} \in \mathcal{X}$ such that $\mu_{\widehat{x}} \geq \max_{x \in \mathcal{X}} \mu_x - \epsilon$ with probability at least $1 - \delta$.

To aid us in our objectives, we assume some structure on the reward function μ.

Assumption 1. *There exists a known feature map $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ that maps each $x \in \mathcal{X}$ to a (possibly infinite dimensional) Hilbert space $\mathcal{H}$, and moreover, there exists a $\theta_* \in \mathcal{H}$ and $h \geq 0$ such that $\max_{x \in \mathcal{X}} |\mu_x - \langle \theta_*, \phi(x) \rangle_{\mathcal{H}}| \leq h$.*

Consequently, if h is not too big, the expected value of each of the observations y_t is nearly a linear function of its associated features $\phi(x_t)$. We say the model is *misspecified* when $h > 0$, and otherwise the setting is well-specified and reduces to the classical stochastic setting when $h = 0$.

Assumption 2. *Rewards are bounded $\max_{x \in \mathcal{X}} |\mu_x| \leq B$.*

Assumption 3. *For every time t, the additive stochastic noise ξ_t is independent, mean-zero with $\mathbb{E}[\xi_t^2] \leq \sigma^2$.*

While we assume the learner knows B and σ^2, we assume that the learner *does not* know the extent of the model misspecification $h \geq 0$. Note that we do *not* assume ξ_t is bounded, indeed, it can even be heavy tailed.

2.1.1 Elimination algorithms and experimental design

Whether the model is misspecified ($h > 0$) or not ($h = 0$), a popular class of algorithms for both the objectives of regret minimization and pure exploration is known as *elimination algorithms*. Elimination algorithms proceed in stages, maintaining a set $\widehat{\mathcal{X}} \subset \mathcal{X}$ of candidates that may achieve $\max_{x \in \mathcal{X}} \mu_x$ given all previous observations. At the beginning of the stage $\ell \geq 1$ the algorithm decides which measurements to take, nature reveals the observations, and the stage ends by constructing an estimate $\widehat{\mu}_{(\cdot)}$ of $\mu_{(\cdot)}$ and removing all elements $x \in \widehat{\mathcal{X}}$ from $\widehat{\mathcal{X}}$ where $\max_{x' \in \widehat{\mathcal{X}}} \widehat{\mu}_{x'} - \widehat{\mu}_x > \epsilon_\ell$. This process is repeated indefinitely in the case of regret minimization, or until $\widehat{\mathcal{X}}$ contains a single element in the case of pure exploration. To be as effective as possible at discarding as many candidates as possible in the elimination stage (without discarding the best arm), a natural strategy of selecting how many and which measurements to take in the beginning of the round is to select $x_1, \ldots, x_n \in \mathcal{X}$ to accurately estimate the differences of the estimates

$$\max_{x, x' \in \widehat{\mathcal{X}}} (\widehat{\mu}_{x'} - \widehat{\mu}_x) - (\mu_{x'} - \mu_x) \leq \epsilon_\ell. \tag{2.1}$$

If $x_* := \arg\max_{x \in \mathcal{X}} \mu_x$ and $x_* \in \widehat{\mathcal{X}}$ at the start of the round, then we have that x_* will not be eliminated at the end since

$$\max_{x' \in \widehat{\mathcal{X}}} \widehat{\mu}_{x'} - \widehat{\mu}_{x_*} \leq \max_{x' \in \widehat{\mathcal{X}}} \mu_{x'} - \mu_{x_*} + \epsilon_\ell \leq \epsilon_\ell.$$

And moreover, it is straightforward to show that after the discarding step of stage ℓ, $\max_{x \in \widehat{\mathcal{X}}} \mu_{x_*} - \mu_x \leq 2\epsilon_\ell$. To guide our choice of $x_1, \ldots, x_n \in \mathcal{X}$ to achieve (2.1), we exploit the assumed (nearly) linear model of above.

2.1.2 Optimal experimental design and the problem of rounding continuous designs

This section introduces the method of experimental design with the goal of achieving (2.1) by taking as few total samples as possible. Shortly, we will consider the case when $h > 0$ and ϕ is an arbitrary feature map. But for now, let us make the simplifying assumption that $h = 0$, ϕ is the identity map so that $\mu_x = \langle \theta_*, x \rangle$, and $\xi_t \sim \mathcal{N}(0, \sigma^2)$. Thus, if at time t we select $x_t \in \mathcal{X} \subset \mathbb{R}^d$ we observe $\langle \theta_*, x_t \rangle + \xi_t$. Suppose we observed pairs $\{(x_t, y_t)\}_{t=1}^T$ where each $x_t \in \mathcal{X}$ was chosen independently of any y_s for $s \leq t$. If we wished to achieve (2.1) for $\widehat{\mathcal{X}} \subset \mathcal{X}$ with $\mu_x = \langle x, \theta_* \rangle$, perhaps the most natural way forward would be to compute the least squares estimator $\widehat{\theta}_{LS} = \arg\min_\theta \sum_{t=1}^T (y_t - \langle x_t, \theta \rangle)^2$, and set $\widehat{\mu}_x = \langle \widehat{\theta}_{LS}, x \rangle$. Then (2.1) is equivalent to $\max_{v \in \mathcal{V}} \langle \widehat{\theta}_{LS} - \theta_*, v \rangle \leq \epsilon_\ell$ with $\mathcal{V} = \widehat{\mathcal{X}} - \widehat{\mathcal{X}}$. By a standard sub-Gaussian tail-bound (Lattimore and Szepesvári, 2020), we have with probability at least $1 - \delta$ that for all $v \in \mathcal{V} \subset \mathbb{R}^d$

$$|\langle v, \widehat{\theta}_{LS} - \theta_* \rangle| \leq \|v\|_{(\sum_{t=1}^T x_t x_t^\top)^{-1}} \sqrt{2\sigma^2 \log(2|\mathcal{V}|/\delta)}, \tag{2.2}$$

where we adopt the notation $\|z\|_A = \sqrt{z^\top A z}$ for any $z \in \mathbb{R}^d$ and symmetric semi-definite positive A. Note that this error bound only depends on those x_t measurements that we choose *before* any responses y_t are observed. This allows us to plan, that is, choose the T measurement vectors to minimize the RHS of (2.2). Unfortunately, this minimization problem is known to be NP-hard (Pukelsheim, 2006; Allen-Zhu et al., 2017). As a consequence, approximation algorithms based on the relaxation

$$\bar{\lambda} = \text{argmin}_{\lambda \in \triangle_\mathcal{X}} \max_{v \in \mathcal{V}} v^\top \left(\sum_{x \in \mathcal{X}} \lambda_x x x^\top \right)^{-1} v \tag{2.3}$$

have been proposed. These first solve for $\bar{\lambda}$ and "round" this to a discrete allocation of measurements.

Deterministic rounding Perhaps the simplest scheme is to obtain a solution $\bar{\lambda}$ of (2.3) and then sample $x \in \mathcal{X}$ exactly $\lceil \bar{\lambda}_x T \rceil$ times. In the worst case, this will result in $|\text{support}(\bar{\lambda})|$ additional measurements than the intended T. Caratheodory's theorem provides a polynomial-time algorithm for constructing $\tilde{\lambda} \in \triangle_{\mathcal{X}}$ such that $\sum_{x \in \mathcal{X}} \tilde{\lambda}_x x x^\top = \sum_{x \in \mathcal{X}} \bar{\lambda}_x x x^\top$ and $|\text{support}(\tilde{\lambda})| \leq (d+1)d/2$. However, more sophisticated rounding procedures exist. (Allen-Zhu et al., 2017) inflates the RHS of (2.2) by a constant factor while only requiring that $T = \Omega(d)$. When $\mathcal{V} = \mathcal{X}$, another strategy is to solve the optimization problem (2.3) with a Frank-Wolfe style algorithm that is terminated only after $O(d \log\log(d))$ iterations so that the rounding according to the naive ceiling operation only inflates T by the number of iterations which is $O(d \log\log(d))$ (Todd, 2016).

Stochastic rounding Another basic rounding algorithm simply samples $x_1, \ldots, x_T \sim \bar{\lambda}$. Unfortunately, using the least squares estimator $\widehat{\theta}_{LS}$, we may have that $\sum_{t=1}^T x_t x_t^\top$ deviates dramatically from $T \sum_{x \in \mathcal{X}} \bar{\lambda}_x x x^\top$ for moderate T, thus any guarantees require T to be poly(d) and moreover, performance relies on the spectrum of $\sum_{x \in \mathcal{X}} \bar{\lambda}_x x x^\top$ (Rizk et al., 2020). As a consequence, (Tao et al., 2018) proposed using the inverse propensity score (IPS) estimator $\widehat{\theta}_{IPS} := (\sum_{x \in \mathcal{X}} \bar{\lambda}_x x x^\top)^{-1} (\frac{1}{T} \sum_{t=1}^T x_t y_t)$. From (Tao et al., 2018), with probability at least $1 - \delta$ we have for all $v \in \mathcal{V}$ simultaneously

$$|\langle v, \widehat{\theta}_{IPS} - \theta_* \rangle| \leq \sqrt{\frac{2\sigma^2 \|v\|^2_{A(\bar{\lambda})^{-1}} \log(2|\mathcal{V}|/\delta)}{T}} + \frac{\log(2|\mathcal{V}|/\delta)(1 + \max_{x \in \mathcal{X}} |v^\top A(\bar{\lambda})^{-1} x|)}{T}. \quad (2.4)$$

where $A(\lambda) := \sum_{x \in \mathcal{X}} \lambda_x x x^\top$. The second term of (2.4) accounts for potentially rare but large deviations of size $\max_{x \in \mathcal{X}} |v^\top A(\bar{\lambda})^{-1} x|$. Sadly, this second term is cumbersome in analyses since it can dominate the first term, and it cannot be removed in the worst-case. A final class of algorithms rely on *proportional volume sampling*, or sampling from a determinantal point process (DPP), but are limited to specific optimality criteria (Nikolov et al., 2019; Derezinski et al., 2020).

2.1.3 Main contributions

The main contributions of the work described in this chapter (Camilleri et al., 2021a) include a novel scheme for experimental design and its application to kernel bandits.

- We propose an estimator $\widehat{\theta}_{RIPS}$ that overcomes many of the shortcomings of the prior art reviewed in Section 2.1.2 for $h = 0$ and $\phi \equiv$ identity. For any fixed $\theta_* \in \mathbb{R}^d$, $\mathcal{V} \subset \mathbb{R}^d, \mathcal{X} \subset \mathbb{R}^d$, $\lambda \in \triangle_{\mathcal{X}}$, and $T \in \mathbb{N}$, if T samples are drawn randomly according to λ to construct $\widehat{\theta}_{RIPS}$, then with probability at least $1 - \delta$ we have for all $v \in \mathcal{V}$

$$|\langle v, \widehat{\theta}_{RIPS} - \theta_* \rangle| \leq \|v\|_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}} \sqrt{\frac{c(\sigma^2 + B^2) \log(2|\mathcal{V}|/\delta)}{T}}$$

  for an absolute constant c. Note that our method puts no restrictions on T but matches the ideal discrete allocation of (2.2) up to a constant by realizing that

$$\frac{\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{v \in \mathcal{V}} \|v\|_{(\sum_{x \in \mathcal{X}} T \lambda_x x x^\top)^{-1}}}{\min_{\{x_t\}_{t=1}^T \in \mathcal{X}} \max_{v \in \mathcal{V}} \|v\|_{(\sum_{t=1}^T x_t x_t^\top)^{-1}}} \leq 1.$$

  We also note that we only assume the stochastic noise has bounded variance and do not rule out heavy-tailed distributions. The estimator $\widehat{\theta}_{RIPS}$ is a special case of our more general estimator.

- We extend our estimator to the misspecified setting where $h \geq 0$ and to use feature maps $\phi : \mathbb{R}^d \to \mathcal{H}$ for an RKHS $\mathcal{H}$. When $\mathcal{H}$ can represent a high or even an infinite dimensional space, restrictions on T based on the dimension start to become paramount. For any fixed $\theta_* \in \mathcal{H}$, $\mathcal{V} \subset \mathcal{H}$, $\mathcal{X} \subset \mathbb{R}^d$, $\lambda \in \triangle_{\mathcal{X}}$, $T \in \mathbb{N}$, and $\gamma \geq 0$, if T samples are drawn randomly according to λ to construct $\widehat{\theta}_{RIPS}(\gamma)$, then with probability at least $1 - \delta$ we have for all $v \in \mathcal{V}$

$$|\langle v, \widehat{\theta}_{RIPS}(\gamma) - \theta_* \rangle| \leq \|v\|_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}} \times \left(\sqrt{\gamma}\|\theta_*\|_2 + h + \sqrt{\frac{c(\sigma^2 + B^2)\log(2|\mathcal{V}|/\delta)}{T}} \right).$$

  Note that since $\mathcal{H}$ may be infinite-dimensional, the estimator $\widehat{\theta}_{RIPS}(\gamma)$ is constructed implicitly and is implemented through kernel evaluations only.

- We empirically compare $\widehat{\theta}_{RIPS}(\gamma)$ to the sampling and estimator pairs of Section 2.1.2 and show that $\widehat{\theta}_{RIPS}(\gamma)$ is competitive on both finite dimensional G-optimal design as well as its regularized RKHS variant sometimes called Bayesian experimental design.

- We employ $\widehat{\theta}_{RIPS}(\gamma)$ in a novel elimination style algorithm for kernel bandits. Our regret bounds match state of the art results in the well-specified setting, and are the first linear bounds that we are aware of for the misspecified setting. In addition, we state an instance-dependent pure-exploration result for identifying an ϵ-good arm with probability at least $1 - \delta$ that compares favorably to known lower bounds. One advantage of our algorithm over prior kernel bandits and Bayesian Optimization algorithms (Srinivas et al., 2009; Valko et al., 2013; Frazier, 2018) is that our approach naturally allows for taking batches of pulls per round.

2.2 Robust Inverse Propensity Score (RIPS) estimator

In this section we introduce the $\widehat{\theta}_{RIPS}$ estimator. In finite dimensions, our estimator first constructs $\widehat{\theta}_{IPS}$ but then to avoid the large deviations term of (2.4) applies robust mean estimation on each $\langle v, \theta_* \rangle$ to obtain a $\widehat{\theta}_{RIPS}$ which is consistent with all of these estimates. When we move to an RKHS setting, we add regularization to avoid vacuous bounds and account for the introduced bias. The bias of misspecification is handled similarly. We begin with robust mean estimation.

Definition 1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean $\bar{x}$ and variance ν^2. Let $\delta \in (0, 1)$. We say that $\widehat{\mu}(X_1, \ldots, X_n)$ is a* δ-robust estimator *if there exist universal constants $c_1, c_0 > 0$ such that if $n \geq c_1 \log(1/\delta)$, then with probability at least $1 - \delta$*

$$|\widehat{\mu}(\{X_t\}_{t=1}^n) - \bar{x}| \leq c_0 \sqrt{\frac{\nu^2 \log(1/\delta)}{n}}.$$

Examples of δ-robust estimators include the median-of-means estimator and Catoni's estimator (Lugosi and Mendelson, 2019). This chapter employs the use of the Catoni estimator which satisfies $|\widehat{\mu}(\{X_t\}_{t=1}^n) - \bar{x}| \leq \sqrt{\frac{2\nu^2 \log(1/\delta)}{n - 2\log(1/\delta)}}$ for $n > 2\log(1/\delta)$ which leads to an optimal leading constant as $n \to \infty$. We will use a separate robust mean estimate for each $v \in \mathcal{V}$. In particular, to estimate $\langle v, \theta_* \rangle$ we use $\widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^T)$ where

$$A^{(\gamma)}(\lambda) := \sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I. \tag{2.5}$$

Our RIPS procedure for experimental design in an RKHS is presented in Figure 2.1. It has the following guarantee.

Algorithm 2.1. RIPS for Experimental Designs in an RKHS

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ, regularization $\gamma > 0$, robust mean estimator $\widehat{\mu} : \mathbb{R}^* \to \mathbb{R}$

$$\lambda^* := \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{v \in \mathcal{V}} \|v\|_{\left(\sum_x \lambda_x \phi(x)\phi(x)^\top + \gamma I\right)^{-1}} \tag{2.6}$$

Randomly draw $\widetilde{x}_1, \ldots, \widetilde{x}_\tau$ from $\mathcal{X}$ according to λ^*
Set $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda^*)^{-1}\phi(\widetilde{x}_t)\widetilde{y}_t\}_{t=1}^{\tau})$
Set $\widehat{\theta} := \arg\min_{\theta} \max_{v \in \mathcal{V}} \dfrac{|\langle \theta, v\rangle - W^{(v)}|}{\|v\|_{(\sum_{x \in \mathcal{X}} \lambda_x^* \phi(x)\phi(x)^\top + \gamma I)^{-1}}}$
Return: $\{W^{(v)}\}_{v \in \mathcal{V}}, \widehat{\theta}$

Figure 2.1: In this chapter, we assume each element in $\mathcal{V}$ is a linear combination of $\phi \circ \mathcal{X}$ which makes all quantities well-defined and can be computed using kernel evaluations $k(x, x') := \langle \phi(x), \phi(x')\rangle_{\mathcal{H}}$. Moreover, Equation 2.6 is convex with gradients that can be computed using kernel evaluations (See Section 2.2.3).

Theorem 1. *Fix any finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ and regularization $\gamma > 0$. If the RIPS procedure of Figure 2.1 is run with $\frac{\delta}{|\mathcal{V}|}$-robust mean estimator $\widehat{\mu}(\cdot)$ and if $\tau \geq c_1 \log(|\mathcal{V}|/\delta)$ then with probability at least $1 - \delta$, we have*

$$\max_{v \in \mathcal{V}} \frac{|W^{(v)} - \langle \theta_*, v\rangle|}{\|v\|_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}}} \leq \sqrt{\gamma}\|\theta_*\|_2 + h + c_0\sqrt{\tfrac{(B^2+\sigma^2)}{\tau}\log(2|\mathcal{V}|/\delta)},$$

Moreover, $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^{\tau})$ can be replaced by $\langle \widehat{\theta}, v\rangle$ by multiplying the RHS by a factor of 2.

Proof sketch. Due to the regularization and potential misspecification if $h > 0$, each $v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t$ is biased. Thus, we apply the guarantee of $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^{\tau})$ to the expectation of its arguments. The triangle inequality followed by repeated applications of Cauchy-Schwartz yields

$$\begin{aligned}
|W^{(v)} - \langle v, \theta_*\rangle| \leq& |W^{(v)} - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1]| + |\mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1] - \langle v, \theta_*\rangle| \\
\leq& c_0\sqrt{\frac{\nu^2 \log(1/\delta)}{\tau}} + \sqrt{\gamma}\|\theta_*\|_2 + h
\end{aligned}$$

where we obtain an upper bound on the variance ν^2 by

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1) &\leq \mathbb{E}[(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1)^2] \\
&= \mathbb{E}\left[\left(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\right)^2 \mu_{x_1}^2\right] + \mathbb{E}\left[\left(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\right)^2 \xi_1^2\right] \\
&\leq (B^2 + \sigma^2)\|v\|^2_{A^{(\gamma)}(\lambda)^{-1}}.
\end{aligned}$$

□

2.2.1 Practical implementation of the algorithms

The construction of $\widehat{\theta}$ in the algorithms may–at first glance–look confusing in the infinite dimensional case. In actuality, the equivalent dual representation $\widehat{\theta} = \sum_{i=1}^{|\mathcal{X}|} \alpha_i \phi(x_i)$ would be used. That is, the potentially infinite dimensional object $\widehat{\theta}$ is represented by a finite dimensional weight vector $\alpha \in \mathbb{R}^{|\mathcal{X}|}$. With that, the optimizations in the algorithms (e.g., to compute the RIPS estimator) are over the dual vector $\alpha \in \mathbb{R}^{|\mathcal{X}|}$, and inner products $\langle \widehat{\theta}, v \rangle = \sum_{i=1}^{|\mathcal{X}|} \alpha_i \langle \phi(x_i), v \rangle$ are computed using the kernel matrix of $\mathcal{X}$ since in all instances of v used in the algorithms, v is a linear combination of $\{\phi(x)\}_{x \in \mathcal{X}}$.

2.2.2 Comparison to IPS estimator

Note the difference between the bound of RIPS in Theorem 1 with the bound of the IPS estimator stated in equation (2.4). Consider the setting of equation (2.4). Ignoring log factors and constants, the confidence bound of the IPS estimator essentially scales as $\sqrt{\frac{\sigma^2 \|v\|^2_{A(\bar{\lambda})^{-1}}}{T}} + \frac{\max_{x \in \mathcal{X}} |v^\top A(\bar{\lambda})^{-1} x|}{T}$, while the confidence bound of RIPS essentially scales as $\sqrt{\frac{\sigma^2 \|v\|^2_{A(\bar{\lambda})^{-1}}}{T}}$. It can be shown that in the instance in the experiment corresponding to figure 2.2c, the term $\frac{\max_{x \in \mathcal{X}} |v^\top A(\bar{\lambda})^{-1} x|}{T} \approx \frac{d}{T}$ while $\sqrt{\frac{\sigma^2 \|v\|^2_{A(\bar{\lambda})^{-1}}}{T}} \approx \frac{\sqrt{d}}{\sqrt{T}}$. Thus, the first term dominates by a polynomial factor in the dimension until $T \geq d$, and the experiment shows that indeed the IPS estimator has larger deviations than RIPS, as suggested by the above upper bounds.

2.2.3 Experimental Design optimization in an RKHS

We now discuss how to actually compute an allocation in a potentially infinite dimensional RKHS $\mathcal{H}$. The following lemma will be helpful and is proved in the appendix.

Lemma 1. *If $A_\lambda = \sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top$ then for $a, b \in \mathcal{H}$*

$$a^\top (A_\lambda + \gamma I)^{-1} b = \frac{1}{\gamma} a^\top b - \frac{1}{\gamma} k_\lambda(a)^\top (K_\lambda + \gamma I_{|\mathcal{X}|})^{-1} k_\lambda(b)$$

with $k_\lambda(\cdot) \in \mathbb{R}^{|\mathcal{X}|}$ so that for any $c \in \mathcal{H}$, $[k_\lambda(c)]_i = \sqrt{\lambda_i} \phi(x_i)^\top c$, and $K_\lambda \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ so that

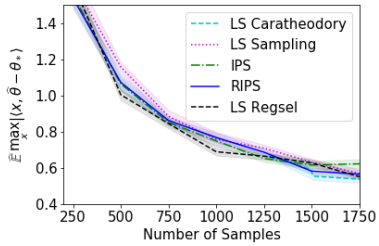
$$[K_\lambda]_{i,j} = \sqrt{\lambda_i} \sqrt{\lambda_j} \phi(x_j)^\top \phi(x_j) =: \sqrt{\lambda_i} \sqrt{\lambda_j} k(x_i, y_i).$$

For $x \in \mathcal{X}$, $[k_\lambda(x)]_i = \sqrt{\lambda_i} \phi(x_i)^\top \phi(x) = \sqrt{\lambda_i} k(x_i, x)$.

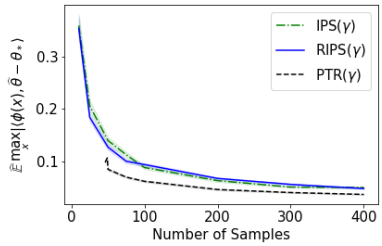
If we call $f(\lambda)$ the argument of Equation 2.6 in Figure 2.1, and $\bar{v} \in \arg\max_{v \in \mathcal{V}} v^\top \left(\sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top + \gamma I \right)^{-1} v$ then the computation of the gradient of $\lambda \mapsto f(\lambda)$ equals

$$[\nabla_\lambda f(\lambda)]_i = -\big(\bar{v}^\top (\sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top + \gamma I)^{-1} \phi(x_i) \big)^2.$$

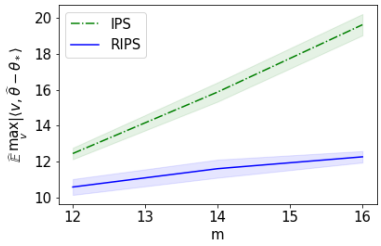
Importantly, in this chapter $\mathcal{V}$ will always be a linear combination of $\{\phi(x)\}_{x \in \mathcal{X}}$ (e.g. $\mathcal{V} = \mathcal{X} - \mathcal{X}$), thus the last quantity can be computed only using kernel evaluations thanks to Lemma 1. We use first order optimization methods to minimize $\lambda \mapsto f(\lambda)$ since it is convex.



(a) G-Optimal Experiment



(b) Kernel Experiment



(c) RIPS vs IPS Experiment

2.2.4 Project-Then-Round (PTR) for RKHS designs

To the best of our knowledge, the RIPS procedure of Figure 2.1 is novel and should be benchmarked. To design a baseline, we take inspiration from previous works on experimental design in an RKHS. For instance, (Alaoui and Mahoney, 2014) employ a sampling distribution related to statistical leverage scores to construct a sketch of the kernel matrix using a Nystrom approximation. The objective in that problem is closest to V-optimal design which aims to minimize the sum-squared error $\sum_{x\in\mathcal{X}} \mathbb{E}[\langle x, \widehat{\theta} - \theta_*\rangle^2]$ (note, our work here is concerned with G-optimal-like objectives, or worst-case error over $\mathcal{X}$). The Nystrom approximation to the kernel matrix effectively projects the problem to a low dimensional sub-space where finite-dimensional rounding techniques like those reviewed in Section 2.1.2 can be applied. (Bach, 2015) also relies on a sampling distribution to approximate integrals using kernels with an objective similar to V optimal.

We describe in Algorithm 2.2 the baseline procedure we call Project-Then-Round (PTR), that employs the finite rounding technique of (Allen-Zhu et al., 2017) described in Section 2.1.2.

Algorithm 2.2. PTR for Experimental Designs in an RKHS

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, regularization $\gamma > 0$
Fix any $\lambda \in \triangle_{\mathcal{X}}$.
Compute $[K]_{i,j} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j)\rangle_{\mathcal{H}}$ the kernel matrix of the set of points $\mathcal{X} = \{x_1, \ldots, x_n\}$.
Consider a decomposition $K = \widehat{\Phi}\widehat{\Phi}^\top$ with $\widehat{\Phi} \in \mathbb{R}^{n\times n}$ such that the rows of $\widehat{\Phi}$ called $\widehat{\phi}(x_i) \in \mathbb{R}^n$ are used to compute $\widehat{A}(\lambda) = \sum_{i=1}^n \lambda_i \widehat{\phi}(x_i)\widehat{\phi}(x_i)^\top$.
Diagonalize $\widehat{A}(\lambda)$ as $\widehat{A}(\lambda) = VDV^\top$ with D diagonal matrix with coefficients $(d_1 \geq d_2 \geq \ldots \geq d_n)$.
Define the *effective dimension* as

$$\widetilde{d}(\lambda, \gamma) = \max\{i \in [n] : d_i \geq \gamma\}.$$

Choose $k = \widetilde{d}(\gamma, \lambda) \in [n]$ and denote V_k as the top k eigenvectors of $\widehat{A}(\lambda)$.
Compute the projections $V_k^\top \widehat{\phi}(x_1), \ldots, V_k^\top \widehat{\phi}(x_n) \in \mathbb{R}^k$
Use the rounding procedure of (Allen-Zhu et al., 2017) to obtain the desired sparse allocation $\{\widetilde{x}_i\}_{i=1}^\tau$.
Return: $\{\widetilde{x}_i\}_{i=1}^\tau$

This procedure enjoys the following guarantees.

Theorem 2. *Consider the procedure of Algorithm 2.2. If the number of measurements τ satisfies $\tau = \Omega(\widetilde{d}(\gamma, \lambda))$, then*

$$\max_{v\in\mathcal{V}} \|v\|^2_{\left(\sum_{i=1}^\tau \phi(\widetilde{x}_i)\phi(\widetilde{x}_i)^\top + \tau\gamma I\right)^{-1}} \leq \max(2, 1+\epsilon) \max_{v\in\mathcal{V}} \|v\|^2_{\left(\sum_{x\in\mathcal{X}} \tau\lambda_x \phi(x)\phi(x)^\top + \tau\gamma I\right)^{-1}}.$$

where $\widetilde{d}(\gamma, \lambda)$ is defined in the algorithm.

We refer the reader to the appendix for the proof of Theorem 2. This procedure performs rounding in a finite dimensional subspace which is a projection of the initial feature space of potentially infinite dimension. With Theorem 2 one can obtain a guarantee similar to that of Theorem 1 up to a constant whenever $\tau = \Omega(\widetilde{d}(\lambda, \gamma))$. Though this effective dimension is rarely the dominating factor in analyses, it is cumbersome to keep around and bound.

2.2.5 Empirical evaluation of allocation methods

We briefly describe illustrative experiments (see the supplementary material for more details).

G-optimal design experiment: We generate $x_1, \ldots, x_n$ by sampling $\tilde{x}_i \sim N(0, \Sigma)$ with $\Sigma_{i,i} = 1$ if $i \leq d - 10$, $\Sigma_{i,i} = .1$ if $i > d - 10$ and all other entries of Σ set to 0. Then, we set $x_i = \frac{\tilde{x}_i}{\|\tilde{x}_i\|}$. We use $\theta_* = \frac{1}{\sqrt{d}}\mathbf{1}$. We set $d = 50$ and $n = \frac{d(d+1)}{2}$. We use mirror descent to solve the G-optimal design problem. We compare RIPS with IPS, Caratheodory's algorithm with the ceiling rounding technique (LS Caratheodory), the rounding technique in (Allen-Zhu et al., 2017) (LS Regsel), and the random sampling approach taken in (Rizk et al., 2020) (LS Sampling). Figure 2.2a depicts the results, and shows that RIPS performs comparably to these other approaches. It also illustrates the shortcomings of the Caratheodory rounding algorithm, which does not return an estimate for $T \leq 1275$, while the other algorithms have already learned nontrivial estimates of θ_* for much smaller values of T.

G-optimal design in an RKHS: We let $\mathcal{X} = \{0, (\frac{1}{m})^2, \ldots, (\frac{m-1}{m})^2, 1\}$ with $m = 500$ and use the RBF kernel $K(x, x') = \exp(-\frac{\|x-x'\|^2}{2\varphi^2})$ with bandwidth parameter $\varphi = 0.025$. Due to this being an infinite dimensional kernel, the ambient dimension for m points is equal to m. We focus on the regime $T < m$ where standard rounding schemes do not apply and compare PTR with regularization γ, $\widehat{\theta}_{RIPS}(\gamma)$, and $\widehat{\theta}_{IPS(\gamma)} := A^{(\gamma)}(\lambda)^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t y_t)$ where we set $\gamma = 0.005$. Figure 2.2b depicts the results, showing that PTR(γ) does slightly better than IPS(γ) and RIPS(γ), and that all three algorithms have learned non-trivial estimates of θ_* using hundreds of samples fewer than standard rounding algorithms require to even output an estimate.

RIPS vs. IPS: While IPS has similar performance to RIPS in the two previous experiments, RIPS performs dramatically better in some settings. Let $m \in \mathbb{N}$ and $d = m^2 + m$. Inspired by combinatorial bandits, we consider a setting where the measurement vectors $\mathcal{X} = \{e_1, \ldots, e_d\}$ consist of the standard basis vectors, $\theta_* = -\mathbf{1}$, and the performance metric for an estimator $\widehat{\theta}$ is $\mathbb{E}\sup_{i\in[m^2]} |v_i^\top(\widehat{\theta} - \theta_*)|$ where $v_i = \sum_{j=1}^{m} e_j + e_{i+m}$. We compare the performance of IPS against RIPS for $m \in \{12, 14, 16\}$ and estimate the expected maximum deviation at $T = 4m$. Figure 2.2c shows that as m grows, the performance of IPS degrades relative to RIPS, reflecting that IPS has large deviations in comparison to our proposed estimator RIPS.

2.3 Algorithms for Kernelized Bandits

We now leverage our proposed RIPS estimator of Algorithm 2.1 for the kernel bandits problem in an elimination style algorithm as introduced in Section 2.1.1. In this section we provide different algorithms to solve the regret minimization and pure exploration problems. This section illustrates the benefits of using our RIPS estimator. In particular, the estimator enables us to design a regret minimization algorithm that trivially supports batching while enjoying state of the art performance in the well-specified setting. In addition, this same algorithm is robust to model misspecification, suffering only linear regret with respect to that

approximation error without any prior knowledge on this error (guarantees that, to the best of our knowledge, our novel). Last but not least, applying our RIPS estimator to pure exploration tasks leads to the first best arm identification provably robust to misspecification.

2.3.1 RIPS for Regret minimization

As introduced in Section 2.1, our objective is to develop an algorithm that minimizes regret under the general stochastic and misspecified setting (Assumptions 1-3). Specifically, when pulling arm $x \in \mathcal{X}$ at time t we observe a random variable $\mu_x + \xi_t$ where ξ_t is independent, mean-zero noise with variance σ^2. We assume there exists a $\theta_* \in \mathcal{H}$ and known feature map $\phi : \mathcal{X} \to \mathcal{H}$ such that $\max_{x \in \mathcal{X}} |\mu_x - \langle \theta_*, \phi(x) \rangle| \leq h$ where $h \geq 0$ is unknown to the learner. That is, μ_x is well-approximated by the linear function $\langle \theta_*, \phi(x) \rangle$ but may deviate from it by an amount $h \geq 0$. Because of model misspecification in the case when $h > 0$, we should not hope to obtain sub-linear regret if we seek a regret bound that grows only logarithmically in $|\mathcal{X}|$ and polynomial in d.

Algorithm 2.3 is a phased elimination strategy where at each round a (regularized) G-optimal design is performed to minimize the variances of the estimates of all the arms and then arms are discarded if their sub-optimality gap is deemed too large (under the assumed linear model). Due to model misspecification, we should only expect this approach to work until hitting a kind of noise floor defined by the level of misspecification h, as suggested from the guarantee from Theorem 1. The algorithm is a combination of our RIPS estimator for the RKHS setting and the robust algorithm of (Lattimore et al., 2020).

Theorem 3. *With probability at least $1 - \delta$, the regret of Algorithm 2.3 satisfies*

$$\sum_{t=1}^{T} \mu_x - \mu_{x_t} \lesssim c_1 \log(|\mathcal{X}|/\delta) + \sqrt{\max_{\mathcal{V} \subset \mathcal{X}} f(\mathcal{V}, \gamma)} \Big(T(h + \sqrt{\gamma}\|\theta_*\|) + \sqrt{c_0^2(\sigma^2 + B^2) T \log(|\mathcal{X}| \log(T)/\delta)} \Big) \tag{2.7}$$

where $f(\mathcal{V}, \gamma) = \inf_{\lambda \in \triangle_{\mathcal{V}}} \sup_{y \in \mathcal{V}} \|\phi(y)\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}}$.

Choosing $\gamma = 1/T$, $\delta = 1/T$ yields an expected regret of

$$\mathbb{E}\big[\sum_{t=1}^{T} \mu_{x_*} - \mu_{x_t}\big] \leq c' \sqrt{\max_{\mathcal{V} \subset \mathcal{X}} f(\mathcal{V}, \tfrac{1}{T})} \big(hT + \sqrt{\log(|\mathcal{X}|T)T} \big)$$

where $c' = O(\sqrt{\|\theta_*\|^2 + \sigma^2 + B^2})$. Note that the hT term due to model misspecification is comparable to the one in (Lattimore et al., 2020). Prior works such as (Srinivas et al., 2009; Valko et al., 2013) have demonstrated expected regret bounds in the well-specified ($h = 0$) setting that scale like $\sqrt{\gamma_T T \log(|\mathcal{X}|)}$ where

$$\gamma_T := \max_{\lambda \in \triangle_{\mathcal{X}}} \log\det(TA^{(0)}(\lambda) + \gamma I). \tag{2.8}$$

where $A^{(0)}(\lambda)$ is defined as in (2.5). The following lemma shows that our own regret bound is never worse than these results.

Lemma 2. *Let γ_T be defined as in* (2.8)*. Then*

$$\max_{\mathcal{V} \subset \mathcal{X}} f(\mathcal{V}, \tfrac{1}{T}) = \max_{\mathcal{V} \subset \mathcal{X}} \inf_{\lambda \in \triangle_{\mathcal{V}}} \sup_{y \in \mathcal{V}} \|\phi(y)\|^2_{A^{(1/T)}(\lambda)^{-1}} \leq \frac{3}{2}\gamma_T \ .$$

Algorithm 2.3. RIPS for Regret Minimization

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ ($|\mathcal{X}| = n$), feature map ϕ, confidence level $\delta \in (0,1)$, regularization γ, sub-Gaussian parameter σ, bound on maximum reward B.

Set $\mathcal{X}_1 \leftarrow \mathcal{X}, \ell \leftarrow 1$

while $|\mathcal{X}_\ell| > 1$ **do**

Let $\lambda_\ell \in \triangle_{\mathcal{X}}$ be a minimizer of $f(\lambda; \mathcal{X}_\ell, \gamma)$ where

$$\begin{aligned} f(\mathcal{V}, \gamma) &= \inf_{\lambda \in \triangle_{\mathcal{V}}} f(\lambda; \mathcal{V}, \gamma) \\ &= \inf_{\lambda \in \triangle_{\mathcal{V}}} \max_{y \in \mathcal{V}} \|\phi(y)\|^2_{(\sum_{y \in \mathcal{V}} \lambda_y \phi(y)\phi(y)^\top + \gamma I)^{-1}} \end{aligned}$$

Set $\epsilon_\ell \leftarrow 2^{-\ell}$, $q_\ell^{(1)} \leftarrow c_1 \log(|\mathcal{X}|/\delta)$

Set $q_\ell^{(2)} \leftarrow c_0^2 (B^2 + \sigma^2) \epsilon_\ell^{-2} f(\mathcal{X}_\ell, \gamma) \log(4\ell^2 |\mathcal{X}|/\delta)$

Set $\tau_\ell \leftarrow \left\lceil \max\left\{ q_\ell^{(1)}, q_\ell^{(2)} \right\} \right\rceil$

Use Algorithm 2.1 with sets $\mathcal{X}_\ell$, $\mathcal{V}_\ell = \phi \circ \mathcal{X}_\ell$, sampling τ_ℓ measurements $x_1, \ldots, x_{\tau_\ell}$ to get $\{W^{(v)}\}_{v \in \mathcal{V}_\ell}$.

Set $\widehat{\theta}_\ell := \arg\min_\theta \max_{v \in \mathcal{V}_\ell} \dfrac{|\langle \theta, v \rangle - W^{(v)}|}{\|v\|_{(\sum_{x \in \mathcal{X}} \lambda_{\ell,x} \phi(x)\phi(x)^\top + \gamma I)^{-1}}}$

Update active set:

$$\mathcal{X}_{\ell+1} = \left\{ x \in \mathcal{X}_\ell, \max_{x' \in \mathcal{X}_\ell} \langle \phi(x') - \phi(x), \widehat{\theta}_\ell \rangle < 4\epsilon_\ell \right\}$$

$\ell \leftarrow \ell + 1$

Play unique element of $\mathcal{X}_\ell$ indefinitely.

The quantity $f(\mathcal{X}, \gamma)$ can also be bounded by a more interpretable form:

Lemma 3. *If* $f(\mathcal{X}, \gamma) = \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{X}} \|\phi(y)\|^2_{(A^{(0)}(\lambda) + \gamma I)^{-1}}$ *then*

$$f(\mathcal{X}, \gamma) \le \mathit{Tr}\left(A^{(0)}(\lambda_D^*)(A^{(0)}(\lambda_D^*) + \gamma I)^{-1} \right) = \mathit{Tr}\left(K_{\lambda_D^*}(K_{\lambda_D^*} + \gamma I)^{-1} \right)$$

where $\lambda_D^* \in \arg\max_{\lambda \in \triangle_{\mathcal{X}}} \log\det\left(A^{(\gamma)}(\lambda)\right)$.

Notably, the RHS of Lemma 3 is the notion of effective dimension that appears in (Alaoui and Mahoney, 2014; Derezinski et al., 2020).

2.3.2 RIPS for Pure Exploration

We consider a slight generalization of the pure exploration setting introduced in Section 2.1. Fix finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Z} \subset \mathbb{R}^d$. We may have $\mathcal{X} = \mathcal{Z}$ but there are interesting cases in which $\mathcal{X} \neq \mathcal{Z}$

Algorithm 2.4. RIPS for Pure Exploration

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Z} \subset \mathbb{R}^d$, feature map ϕ, confidence level $\delta \in (0,1)$, regularization γ, sub-Gaussian parameter σ, bound on maximum reward B, bound on the misspecification noise h.

Let $\mathcal{Z}_1 \leftarrow \mathcal{Z}, \ell \leftarrow 1$

while $|\mathcal{Z}_\ell| > 1$ **do**

Let $\lambda_\ell \in \triangle_{\mathcal{X}}$ be a minimizer of $f(\lambda; \mathcal{Z}_\ell; \gamma)$ where

$$\begin{aligned} f(\mathcal{V}; \gamma) &= \inf_{\lambda \in \triangle_{\mathcal{X}}} f(\lambda; \mathcal{V}; \gamma) \\ &= \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{v, v' \in \mathcal{V}} \|\phi(v) - \phi(v')\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}} \end{aligned}$$

Set $\epsilon_\ell \leftarrow 2^{-\ell}$, $q_\ell^{(1)} \leftarrow c_1 \log(|\mathcal{Z}|/\delta)$

Set $q_\ell^{(2)} \leftarrow c_0^2 \epsilon_\ell^{-2} f(\mathcal{Z}_\ell; \gamma)(B^2 + \sigma^2) \log(2\ell^2 |\mathcal{Z}|^2/\delta)$

Set $\tau_\ell \leftarrow \left\lceil \max\left\{ q_\ell^{(1)}, q_\ell^{(2)} \right\} \right\rceil$

Use Algorithm 2.1 with sets $\mathcal{X}$, $\mathcal{V}_\ell = \phi \circ \mathcal{Z}_\ell - \phi \circ \mathcal{Z}_\ell$, sampling τ_ℓ measurements $x_1, \ldots, x_{\tau_\ell}$ to get $\{W^{(v)}\}_{v \in \mathcal{V}_\ell}$.

Set $\widehat{\theta}_\ell := \arg\min_\theta \max_{v \in \mathcal{V}_\ell} \dfrac{|\langle \theta, v \rangle - W^{(v)}|}{\|v\|_{(\sum_{x \in \mathcal{X}} \lambda_{\ell,x} \phi(x)\phi(x)^\top + \gamma I)^{-1}}}$

$\mathcal{Z}_{\ell+1} = \left\{ z \in \mathcal{Z}_\ell : \max_{z' \in \mathcal{Z}_\ell} \langle \phi(z') - \phi(z), \widehat{\theta}_\ell \rangle \le 2\epsilon_\ell \right\}$

$\ell \leftarrow \ell + 1$

Output: $\mathcal{Z}_\ell$

including combinatorial bandits and recommendation tasks (Fiez et al., 2019). We say a $z \in \mathcal{Z}$ is ϵ-good if $\mu_z \ge \max_{z' \in \mathcal{Z}} \mu_{z'} - \epsilon$. In the pure exploration game, for $\epsilon > 0$ and $\delta \in (0,1)$ the player seeks to identify an ϵ-good arm by taking as few measurements in $\mathcal{X}$ as possible. Just as in regret minimization games, we assume that when the player at time t plays $x_t \in \mathcal{X}$ she observes $y_t = \mu_{x_t} + \xi_t$ where ξ_t is independent mean-zero noise with variance σ^2. Finally, we assume the existence of a $\theta_* \in \mathcal{H}$ such that

$$\max\left\{ \max_{z \in \mathcal{Z}} |\mu_z - \langle \theta_*, \phi(z) \rangle|, \max_{x \in \mathcal{X}} |\mu_x - \langle \theta_*, \phi(x) \rangle| \right\} \le h$$

for some $h \ge 0$ that is *unknown* to the player.

Consider the elimination style algorithm of Algorithm 2.4. The algorithm is a combination of our RIPS procedure and the algorithm of (Fiez et al., 2019). While the algorithm is inspired by (Fiez et al., 2019), their analysis only holds in the well-specified setting ($h = 0$), hence a new proof technique was necessary to achieve the following result for general $h \ge 0$.

Theorem 4. *With $z_* \in \arg\max_{z\in\mathcal{Z}} \langle z, \theta_*\rangle$, fix any $\epsilon \geq \bar{\epsilon}$ where*

$$\bar{\epsilon} = 8\min\{\epsilon \geq 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{g(\epsilon)}) \leq \epsilon\},$$
$$g(\epsilon) = \inf_{\lambda\in\triangle_\mathcal{X}} \sup_{z\in\mathcal{Z}:\langle\theta_*,\phi(z_*)-\phi(z)\rangle\leq\epsilon} \|\phi(z_*) - \phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}}$$

Then with probability at least $1-\delta$, once the algorithm has taken at least τ samples where $\tau = \widetilde{O}(c_1 \log(|\mathcal{Z}|/\delta) + \log(\epsilon^{-1})c_0^2(B^2+\sigma^2)\log(|\mathcal{Z}|/\delta)\rho^*(\gamma,\epsilon))$ we have that $\mu_{\widehat{z}} \geq \max_{z'\in\mathcal{Z}} -\epsilon$ where $\widehat{z}$ is any arm in the set $\mathcal{Z}_\ell$ under consideration after τ pulls and

$$\rho^*(\gamma,\epsilon) = \inf_{\lambda\in\Delta_\mathcal{X}} \sup_{z\in\mathcal{Z}} \frac{\|\phi(z_*) - \phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{\max\{\epsilon^2, \langle\theta_*, \phi(z_*) - \phi(z)\rangle^2)\}}. \tag{2.9}$$

Note that if $\mathcal{X} = \mathcal{Z}$ we have

$$\begin{aligned} g(\epsilon) &= \inf_{\lambda\in\triangle_\mathcal{X}} \sup_{z\in\mathcal{X}:\langle\theta_*,\phi(z_*)-\phi(z)\rangle\leq\epsilon} \|\phi(z_*) - \phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}} \\ &\leq 4 \inf_{\lambda\in\triangle_\mathcal{X}} \sup_{x\in\mathcal{X}} \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}} \\ &\leq 4\mathrm{Tr}((A^{(0)}(\lambda^*_D) + \gamma I)^{-1} A^{(0)}(\lambda^*_D)) \end{aligned}$$

where the last line follows from Lemma 3. This means $\bar{\epsilon}$, the limit on how well one can estimate the maximizing arm, satisfies $\bar{\epsilon} \lesssim (\gamma\|\theta_*\| + h)\mathrm{Tr}((A^{(0)}(\lambda^*_D) + \gamma I)^{-1} A^{(0)}(\lambda^*_D))^{1/2}$. Thus, if we seek an ϵ-good arm, we should choose γ to make this right hand side less than ϵ. Note that $\gamma = 0$ and $h = 0$ implies $\bar{\epsilon} = 0$. If $\phi \equiv$ identity so that $\mathcal{H} = \mathbb{R}^d$, $h = 0$, and $\gamma = 0$ then the sample complexity of Theorem 4 is known to be optimal up to log factors to identify the very best arm (assuming it is unique) relative to any δ-correct algorithm over $\theta_* \in \mathbb{R}^d$ (Soare et al., 2014; Fiez et al., 2019).

2.3.3 Comparing to the alternative baseline procedure

In Section 2.2.4 we proposed a natural alternative to our RIPS procedure for experimental design in an RKHS. This PTR baseline leveraged the fact that the added regularization $\gamma > 0$ effectively made many directions irrelevant. Thus, it projected the problem to a low dimensional subspace where it could apply any of the standard rounding techniques for finite dimensions described in Section 2.1.2. The dimension of this subspace, denoted $\widetilde{d}$, scales like the number of eigenvalues of $\sum_{x\in\mathcal{X}} \lambda^*_x \phi(x)\phi(x)^\top$ that are greater than γ where $\lambda^* \in \arg\min_{\lambda\in\triangle_\mathcal{X}} \max_{v\in\mathcal{V}} \|v\|^2_{\left(\sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I\right)^{-1}}$. Any standard rounding algorithm would then require the number of samples taken from the design to be at least $\widetilde{d}$. Relative to our results, this inflates our regret bound and sample complexity by an additive factor of $\widetilde{d}$ scaled by some problem-dependent $\log$ factors. Algebra shows that $\widetilde{d} \leq 2\mathrm{Tr}((A(\lambda^*_D) + \gamma I)^{-1} A(\lambda^*_D))$. Though for regret this is a lower order term, for pure-exploration with $\mathcal{X} \neq \mathcal{Z}$, this term may potentially dominate the sample complexity because it does not capture the interplay between the geometry of $\mathcal{X}$ and $\mathcal{Z}$. Fortunately, our RIPS procedure demonstrates it is unnecessary and avoids it.

2.4 Related work

There exist excellent surveys of experimental design from both a statistical and computational perspective (Pukelsheim, 2006; Atkinson et al., 2007; Todd, 2016). This chapter is particularly interested in the task

of converting a continuous design into a discrete allocation of T measurements. We reviewed a number of works in Section 2.1.2 for completing this task in finite dimensions. To move to an RKHS setting we considered a regularized design objective which is also known as Bayesian experimental design (Chaloner and Verdinelli, 1995; Allen-Zhu et al., 2017; Derezinski et al., 2020). While most Bayesian experimental design works assume a low-dimensional ambient space and use simple rounding, one exception is the work of (Alaoui and Mahoney, 2014) that performs experimental design in an RKHS for a different design objective, which inspired our project-then-round procedure described of Section 2.2.4. And very recently, (Derezinski et al., 2020) proposed a method of sampling from a determinantal point process (DPP) and showed that they can approximate many continuous experimental design objectives up to a constant factor if $T \gtrsim d_{eff} := \mathrm{Tr}((A(\lambda_D^*) + \gamma I)^{-1} A(\lambda_D^*))$ with λ_D^* defined in Lemma 3. However, according to Table 1 of (Derezinski et al., 2020) the method may not apply to G-optimal-like objectives[1], which is the primary objective of this chapter. To our knowledge, our proposed RIPS method is novel in that its performance is directly comparable to the continuous design without requiring a minimum number of measurements with some dependence on the (effective) dimension. However, our method does require the number of measurements to exceed $\log(|\mathcal{V}|)$. While we leveraged experimental design techniques for kernel bandits, many prior works were able to obtain regret bounds and pure-exploration results using other methods.

Kernel bandits In the well-specified setting ($h = 0$) (Srinivas et al., 2009) propose a UCB style algorithm (Auer et al., 2002a) for the RKHS setting. Independently, (Grünewälder et al., 2010) developed similar methods for minimizing simple regret. (Srinivas et al., 2009) established a regret bound of $\sqrt{T}(\|\theta_*\|\sqrt{\gamma_T} + \gamma_T)$ where γ_T is defined in (2.8). (Valko et al., 2013) proposed another UCB variant to obtain a regret bound that scales just as $\|\theta_*\|\sqrt{p_T T}$ where p_T is an algorithm-dependent constant that can be upper bounded by γ_T, thus improving (Srinivas et al., 2009). We recall that our own regret bound of Theorem 3 scales no worse than $\|\theta_*\|\sqrt{\gamma_T T}$ using Lemma 2, thus matching state of the art. (Chowdhury and Gopalan, 2017) offer improvements in regret over GP-UCB when the action space is infinite. We also note that our algorithm naturally allows batch querying, a property that UCB-like algorithms achieve only through inelegant means (Desautels et al., 2012; Wu and Frazier, 2018).

Misspecified models Our approach to misspecified models draws inspiration from (Lattimore et al., 2020) which addresses linear bandits in finite dimensions. Their regret bound scales quadratically in the ambient dimension due to rounding effects. Our RIPS procedure extends this work to an RKHS. The misspecified model setting is related to the corrupted setting where an adversary can choose to corrupt the observed reward by c_t in each round t. Any algorithms for this adversarial setting can also be used to solve kernelized multi-armed bandit in the misspecified setting with total amount of corruption equal to at most $C_T = \sum_{t=1}^{T} c_t = hT$. Using this reduction, the regret bound for the corrupted setting of (Bogunovic et al., 2020) scales like $C_T\sqrt{\gamma_T T}$. Unfortunately, if we take $C_T = hT$ this bound is vacuous. Whether robust algorithms like (Gupta et al., 2019) can be extended to our kernel bandit setting is an open question. Concurrently, (Lee et al., 2021) independently proposed a very similar estimator and algorithm for the related task of solving adversarial bandits.

Constrained linear bandits If we assumed that $\|\theta_*\|_2 \leq R$ for some explicit, known $R > 0$ then this setting is known as constrained linear bandits, tackled in (Degenne et al., 2020) for the pure-exploration and (Tirinzoni et al., 2020) for the regret setting, respectively. There, a lower bound on the sample complexity of identifying the best arm can be computed. The lower bound is $\inf_{\lambda \in \Delta_{\mathcal{X}}} \sup_{x' \neq x_*} \inf_{\gamma \geq 0} G^{-1}(\lambda, x, \gamma)$ where

$$G(\lambda, x, \gamma) = \frac{\max\{(x' - x_*)^\top (A(\lambda) + \gamma I)^{-1} A(\lambda)\theta_*, 0\}^2}{2\|x' - x_*\|^2_{(A(\lambda)+\gamma I)^{-1}}} + \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right),$$

[1]Our Theorem 2 with the fact $\widetilde{d} \leq 2d_{eff}$ suggests k only needs to be at least $\widetilde{d}$ for G-like objectives, which adds to their table.

which is close to our upper bound ρ^* from equation 2.9. See Corollary 4 for the proof of this lower bound. (Degenne et al., 2020) propose an algorithm with an asymptotic upper bound in the sense that as $\delta \to 0$, the dominant term matches the lower bound. However, while (Degenne et al., 2020) and (Tirinzoni et al., 2020) are tight asymptotically, they suffer from large sub-optimal dependencies on problem-specific parameters.

2.5 Conclusion

In this chapter, we have brought to the non-parametric learning setting an estimator that relies on continuous designs while enjoying state of the art - theoretical and experimental - guarantees for both the well-specified and the misspecified settings. We leveraged this estimator in a novel elimination style algorithm for kernel bandits. For the most part we have ignored computation. However, the computational cost of the RIPS estimator scales *linearly* in $|\mathcal{V}|$. An interesting avenue of research is designing an estimator that leverages multi-dimensional robust mean estimation that has the same properties as RIPS but has *no* dependence on $|\mathcal{V}|$. Such an estimator would be of considerable interest in problems such as combinatorial bandits where $|\mathcal{V}|$ is potentially exponential in the dimension (e.g., see (Katz-Samuels et al., 2020; Wagenmaker et al., 2021)).

Chapter 3

Nearly Optimal Algorithms for Level Set Estimation

3.1 Introduction

The level-set of a function is a subset of its domain where it exceeds a specific value. Level set estimation is the problem of identifying a subset that approximates the true level-set based on a finite set of potentially noisy function evaluations. As an example, consider the goal of detecting a region in a body of water, such as a channel, that is at least $20m$ deep for ships to safely pass. Given that we can obtain noisy estimates of depth using a sonar device at the locations of our choosing, where should we measure in order to acquire the most accurate level-set estimation while using as few total measurements as possible? Level-set estimation can also be interpreted as a kind of classification rule. For example, using as few total experiments as possible, we may want to identify all compounds among a given finite set under consideration that have some property (e.g., binding affinity) that exceeds some target threshold.

While level-set estimation is somewhat of a well-studied problem, to date there is a lack of theoretical understanding of the limits and tradeoffs of estimation accuracy and number of measurements. Most algorithms proceed by sequentially and greedily optimizing an *acquisition function* that is constructed using all the measurements observed up to the current time. These heuristics are known to work very well in practice, but their guarantees are ad hoc and, at best, worst-case (minimax). In this chapter we are interested in understanding the instance-dependent sample complexity of level-set estimation. That is, we would like for an algorithm to output a satisfactory estimate of the level-set as fast as *any* algorithm could for *this particular* instance, not some worst-case instance.

In contrast to prior works that propose a sampling heuristic–usually based on identifying an informative point–and bound its sample complexity, we work backwards. Namely, we first consider an information theoretic lower bound for the level-set estimation problem that suggests an "optimal" sampling strategy. Because this ideal sampling strategy is a function of the true (unknown) function, it is a priori impossible to realize. Instead, we propose a series of sampling strategies, based on *experimental designs*, that mimic this optimal sampling strategy given the information available at the current time. By the end, these strategies provably achieve the optimal sample complexity with minimal overhead. Furthermore, we show that our sampling strategy leads to an upper bound on the sample complexity that is tighter than those in the existing literature. In what follows, we first formally state the problem and our desired objectives. We then review the related work in context before proceeding to our lower bounds and algorithms. We finish with experiments contrasting with existing work.

3.1.1 Problem Statement

We assume there exists an unknown function $f : \mathbb{R}^d \to [-B, B]$ and a subset of allowable sampling locations $\mathcal{X} \subset \mathbb{R}^d$ which span $\mathbb{R}^d$. Though the function f is unknown, we may query its value for any $x \in \mathcal{X}$ and receive a noisy estimate $f(x) + \eta$ where η is iid, $\mathbb{E}[\eta] = 0$, and $\mathbb{E}[\eta^2] \leq \sigma^2$. We define two objectives.

Explicit Level Set Estimation: Given a specified threshold $\alpha \in \mathbb{R}$, the goal is to identify $G_\alpha := \{x \in \mathcal{X} : f(x) > \alpha\}$.

Implicit Level Set Estimation: Let $x_* \in \arg\max_{x \in \mathcal{X}} f(x)$. Given $\epsilon > 0$, the goal is to identify $G_\epsilon := \{x \in \mathcal{X} : f(x) > (1 - \epsilon) f(x_*)\}$[1].

Consider an algorithm that at each time t selects an arm $x_t \in \mathcal{X}$ that is measurable with respect to a σ-algebra $\mathcal{F}_{t-1} = \sigma(x_1, y_1, \cdots, x_{t-1}, y_{t-1})$ and receives a value $y_t = f(x_t) + \eta_t$. To be precise, we say that an algorithm is *PAC-δ* for the explicit (respectively implicit) level set problem if it stops at a time T_δ which is measurable with respect to the filtration $(\mathcal{F}_t)_{t \geq 1}$ and returns G_α (and in the implicit setting returns G_ϵ) with probability at least $1 - \delta$. If $f(x)$ is very close to the threshold, it may take an enormous number of samples to determine whether it is above or below the threshold, so in practice we introduce a $\widetilde{\beta} \geq 0$ tolerance that ensures that any learner has a finite sample complexity (see theorems) and allows for misclassification of points very near to the threshold. But in the discussion that follows, assume that $f(x)$ is bounded away from the threshold.

Our approach is based on modeling f in a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$. Let $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ be the "feature map" associated with the RKHS. Since $|f(x)| \leq B$ for all $x \in \mathcal{X}$, there exists a $\theta_* \in \mathcal{H}$ and a scalar $h \geq 0$ such that $\max_{x \in \mathcal{X}} |f(x) - \langle \theta_*, \phi(x) \rangle_{\mathcal{H}}| \leq h$. When $h = 0$, $f \in \mathcal{H}$, and in general we allow $h \geq 0$ (typically small) in the interest of generality. Our sample complexity bounds will depend on h and $\|\theta_*\|_{\mathcal{H}}$ which we denote $\|\theta_*\|$. If h is small, then f is well approximated as a linear function of the feature maps $\phi(x)$. We refer to the case when $h > 0$ as being *misspecified* and otherwise when $h = 0$ as being *well-specified*. This class of functions is frequently used for level-set estimation because it is often sufficiently rich to model real-world functions but also contains enough structure to quantify the uncertainty of generalizing a learned function to unmeasured locations. One note of departure from the existing literature is that we do not assume the unknown function is precisely captured by a function in an RKHS, only that it is well approximated by one (i.e., the misspecified setting). In the discussion that follows, we additionally assume $|\mathcal{X}| < \infty$ for simplicity since in practice given an arbitrary bounded domain we can replace $\mathcal{X}$ with a finite cover.

3.2 Related Work

The level-set estimation problem naturally connects to several related ideas in Bayesian optimization and multi-armed bandits. In the former setting, methods tend to sample greedily according to an acquisition function that seeks to minimize the uncertainty of the learner about the level set. The first work on level set estimation that employed the use of Gaussian processes and introduced the Straddle heuristic is due to Bryan et al. (2005). These ideas were further developed in Gotovos (2013) which proposed the LSE and LSE-imp algorithms for explicit and implicit level set respectively. They provide a theoretical guarantee on the sample complexities of LSE and LSE-imp, and as we will show below, our sample complexity is always at least as good as their stated bounds. Bogunovic et al. (2016) further connected Bayesian optimization with level set estimation and considered the setting of heteroscedastic noise. The work of Shekhar and Javidi

[1]For ease of exposition, we assume $f(x_*) \geq 0$. This is easily removed by taking $\epsilon < 0$ if $f(x_*) < 0$.

(2019) focuses on the level-set problem in a continuous domain, and provides an algorithm that maintains a notion of uncertainty over regions, providing a potentially improved computational complexity, along with tighter sample complexity bounds compared to LSE for certain kernels and smoothness assumptions. The work of Zanette et al. (2018) reposes level-set estimation as a classification problem and introduces a novel acquisition function. Iwazaki et al. (2019) extends the work of Zanette et al. (2018) to improve model robustness in quality control applications. Bogunovic (2019); Vakili et al. (2021a) demonstrate frequentist guarantees for Gaussian process algorithms. (Bect et al., 2012; Azzimonti et al., 2021) employ a sequential experimental design approaches for estimating failure probability given a threshold form a density that is expensive to evaluate. (Chevalier et al., 2014) proposes a kriging-based approach for the same problem. This line of work is also related to Gaussian Process Bandits, namely the GPUCB algorithm and improved variants (Srinivas et al., 2009; Chowdhury and Gopalan, 2017; Valko et al., 2013). Ha et al. (2020) introduces a Bayesian Neural Network approach for active level set estimation using Monte Carlo dropout techniques. Table B.1 in the appendix summarizes the results we are aware of in the Gaussian process setting.

In the multi-armed and linear bandit setting, the explicit level set estimation problem is related to threshold bandits where one seeks to find all arms above an explicit threshold (Locatelli et al., 2016; Jamieson and Jain, 2018; Degenne et al., 2020). The approach of Degenne et al. (2020), would provide an asymptotically optimal algorithm in the linear setting, however we are not aware of any other works that provide an optimal finite-time guarantee. The implicit level set problem in the standard multi-armed bandit setting is equivalent to the *multiplicative all-ϵ* problem introduced by Mason et al. (2020). Algorithm 3.2 recovers the sample complexities of the instance-optimal $(\texttt{ST})^2$ algorithm given there. Finally, our experimental design techniques are inspired by Soare et al. (2014); Fiez et al. (2019), and especially our work Camilleri et al. (2021a) – presented in Chapter 2 – that introduces the RIPS estimator which we use to perform experimental design in an RKHS.

3.3 Explicit Level Set Estimation

In recent years, adaptive experimental design has arisen as a popular paradigm for active learning in structured settings, for example in linear bandits and RKHS (Soare et al., 2014; Fiez et al., 2019; Camilleri et al., 2021a), and we adapt these ideas for the level set problem. To motivate this paradigm, in the following example we focus on the well-specified linear case where $\phi(x) = x, \widetilde{\beta} = 0, h = 0$ where we recall h denotes the misspecification and $\widetilde{\beta}$ denotes the error tolerance as defined in Section 3.1.1. Imagine we have access to a collection of n-measurements $\{(x_i, y_i)\}_{i=1}^n$ and let $\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}^d} \sum_{i=1}^n (y_i - x_i^\top \theta)^2$ be the least squares estimator. Standard results show that with probability greater than $1-\delta$, we have for all $x \in \mathcal{X}$ simultaneously

$$|x^\top(\widehat{\theta} - \theta_*)| \leq \|x\|_{\left(\sum_{i=1}^n x_i x_i^\top\right)^{-1}} \sqrt{\frac{2\log(2|\mathcal{X}|/\delta)}{n}},$$

where the additional factor of $|\mathcal{X}|$ in the logarithm arises from a union bound over $\mathcal{X}$. In particular, if our data is chosen so that for each arm $x \in \mathcal{X}$

$$|x^\top \theta_* - \alpha| > \|x\|_{\left(\sum_{i=1}^n x_i x_i^\top\right)^{-1}} \sqrt{\frac{2\log(2|\mathcal{X}|/\delta)}{n}}, \tag{3.1}$$

we see that for any x such that $x^\top \theta_* > \alpha$,

$$x^\top \widehat{\theta} = x^\top \theta_* + x^\top(\widehat{\theta} - \theta_*) > x^\top \theta_* - |x^\top \theta_* - \alpha| > \alpha.$$

The first inequality stems from equation (3.1) where we have sampled such that the error $x^\top(\widehat{\theta} - \theta_*)$ is less that the margin to the threshold $|x^\top\theta_* - \alpha|$. Hence, if $x^\top\theta_* > \alpha$ then $x^\top\widehat{\theta} > \alpha$. This same argument may be repeated for $x : x^\top\theta_* < \alpha$. Therefore $\{x : x^\top\widehat{\theta} > \alpha\} = \{x : x^\top\theta_* > \alpha\} = G_\alpha$, i.e. we have a high probability guarantee that we return the correct set of arms above the threshold. Letting $\lambda_x = n_x/n$ be the proportion of times we sample $x \in \mathcal{X}$, we see see that equation (3.1) is equivalent to

$$n \geq \max_{x \in \mathcal{X}} \frac{\|x\|^2_{\left(\sum_{x \in \mathcal{X}} \lambda_x x x^\top\right)^{-1}}}{(\theta_*^\top x - \alpha)^2}. \tag{3.2}$$

In particular, this implies that to achieve a good sample complexity we can minimize the right side of this expression over all possible distributions $\lambda \in \triangle_\mathcal{X}$ where $\triangle_\mathcal{X} = \{\lambda \in \mathbb{R}^{|\mathcal{X}|} : \sum_{x \in \mathcal{X}} \lambda_x = 1, \lambda_x \geq 0 \; \forall x\}$. Indeed as the following theorem shows, this gives a lower bound on this problem.

Theorem 5. *Assume $\eta_t \overset{iid}{\sim} \mathcal{N}(0,1) \; \forall t$. In the well-specified linear setting when $\phi(x) = x$ and $f(x) = \theta_*^\top x$, for any $\delta > 0$, any PAC-δ algorithm with stopping time T_δ that returns the set G_α with probability at least $1 - \delta$ must satisfy*

$$\frac{\mathbb{E}[T_\delta]}{\log(1/2.4\delta)} \geq 2 \min_{\lambda \in \triangle_\mathcal{X}} \max_{x \in \mathcal{X}} \frac{\|x\|^2_{A(\lambda)^{-1}}}{(\theta_*^\top x - \alpha)^2}$$

where $A(\lambda) := \sum_{x \in \mathcal{X}} \lambda_x x x^\top$.

Remark. We prove this result for completeness in the appendix using ideas from Fiez et al. (2019). A similar result has appeared previously in the Appendix of Degenne et al. (2020) which also shows its tightness.

As a concrete interpretation of the lower bound, consider the case where $x_i = e_i$, the i^{th} standard basis vector. Then the mean of arm i is $\theta_*^\top e_i = [\theta_*]_i$, the i^{th} entry of θ_*. This setting removes all structure by making the mean of each point independent of the others, and we may solve the optimization in Theorem 5 in closed form. Namely, the fraction of samples given to arm i, denoted $\lambda^{(i)} \propto ([\theta_*]_i - \alpha)^{-2}\log(1/\delta)$ is proprotional to its inverse gap squared. This leads to a lower bound of $\mathbb{E}[\tau_\delta] \geq \sum_{i=1}^n ([\theta_*]_i - \alpha)^{-2}\log(1/\delta)$ with matches the known lower bounds from (Jamieson and Jain, 2018; Locatelli et al., 2016) which are specific to this setting.

We now operationalize this lower bound to provide an algorithm for level set estimation that has a nearly matching upper bound. In the following sections, we will explain our algorithm and the adaptations necessary to handle the general setting of the RKHS.

3.3.1 Algorithm

Motivated by this lower bound, we now provide an experimental design approach in the general case. In this setting, we recall the feature map $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ and $h \geq 0$ represents the possibly nonzero misspecification level. Despite these changes, the same intuition from the linear case in Theorem 5 applies. We have a set of vectors $\phi(x)_1, \ldots, \phi(x_n) \in \mathcal{H}$ and an unknown parameter vector $\theta_* \in \mathcal{H}$ such that $f(x) \approx \theta_*^\top\phi(x)$. Ideally, we would sample according to a distribution λ_* that achieves the minimum in the lower bound in Theorem 5, however this is not possible since λ_* depends on the a priori unknown θ_*. Instead, we approximate this distribution by solving a series of designs based on the information we have thus far. Furthermore, we allow for a tolerance $\tilde{\beta} \geq 0$ reflecting the fact that depending on the setting, practitioners may be satisfied with an approximate solution if it requires fewer samples to learn.

Our approach, MELK (Misspecified Explicit Level set via Kernelization), for the generalized RKHS setting is given in Algorithm 3.1. MELK proceeds in phases. To keep track of the points it has identified so far, MELK maintains two sets: 1) $\widehat{G}_t$ is the set of all points that up to round t have been declared as being in G_α by MELK, that is $f(x) > \alpha$. 2) $\widehat{B}_t$ is the set of all points declared as being in G_α^c. The remaining, uncertain points are *active* and in the set $\mathcal{A}_t$. Motivated by the lower bound from the linear setting, it then computes the experimental design: $\lambda_t = \arg\min_{\lambda \in \triangle_\mathcal{X}} \max_{x \in \mathcal{A}_t} \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}$ with $A^{(\gamma)}(\lambda) := \sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I$ where γ is a necessary regularization in the kernelized (infinite-dimensional) setting. Indeed, the number of samples taken in each round equals $N_t \approx \min_\lambda \max_{x \in \mathcal{A}_t} \frac{\|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{(2^{-t})^2}$ from λ_t. This guarantees that at the end of the round, $\mathcal{A}_{t+1} \subset \{x \in \mathcal{X} : |\theta_*^\top x - \alpha| \leq 2^{-(t+1)}\}$ and, we can interpret our design as an approximation to the lower bound on the points that are remaining. MELK declares that $x \in G_\alpha$ if $\widehat{\theta}^T \phi(x) - 2^{-t} \gtrsim \alpha$ and adds x to the set G_t. Similarly, MELK adds x to declares $x \in G_\alpha^c$ and adds x to B_t if $\widehat{\theta}^T \phi(x) + 2^{-t} \lesssim \alpha$. Finally, MELK terminates when either all arms have been added to the sets G_t or B_t or when $t \gtrsim \log_2(1/\widetilde{\beta})$ and it has achieved the practitioner's desired tolerance of $\widetilde{\beta}$.

MELK leverages a Robust Inverse Propensity Scoring (RIPS) estimator introduced in Camilleri et al. (2021a) presented in Chapter 2 and further reviewed in Appendix B.2. Previous works in linear bandits have utilized rounding procedures for sampling followed by ordinary least squares that are not applicable in the infinite dimensional setting. Instead, the RIPS estimator appeals to an inverse propensity score estimator plus robust mean estimation. We remind the guarantee of the RIPS estimator below, as it is important for understanding the behavior of the algorithms we present, but the experimental designs we propose are not consequences of that work.

Theorem 6 (Rephrasing of Theorem 1)**.** *Consider the model $y = \langle \phi(x), \theta_* \rangle_\mathcal{H} + \zeta_x + \eta$ for misspecification $|\zeta_x| \leq h$ where it is assumed that $|\langle \phi(x), \theta_* \rangle_\mathcal{H} + \zeta_x| \leq B$, $\mathbb{E}[\eta] = 0$, and $\mathbb{E}[\eta^2] \leq \sigma^2$. Fix any finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ, regularization $\gamma > 0$, and distribution $\lambda \in \triangle_\mathcal{X}$. If $\tau \geq 2\log(|\mathcal{V}|/\delta)$ then with probability at least $1 - \delta$, RIPS returns $\widehat{\theta}$ satisfying*

$$\max_{v \in \mathcal{V}} \frac{|\langle \widehat{\theta}, v \rangle - \langle \theta_*, v \rangle|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} \leq 2\sqrt{\gamma}\|\theta_*\| + 2h + 4\sqrt{\tfrac{(B^2+\sigma^2)}{\tau} \log\left(\tfrac{2|\mathcal{V}|}{\delta}\right)}.$$

Computational Considerations. We note briefly that while we state the optimal design in terms of the potentially infinite dimensional $\phi(x)$ for clarity, we never explicitly compute $\phi(x)$ and instead resort to the kernel trick (see Appendix B.6). Furthermore the design can be computed using first order optimization methods, such as Frank-Wolfe (Lattimore and Szepesvári, 2020; Todd, 2016). The total computational cost of each design is poly($|\mathcal{X}|$). Though these designs can be expensive to compute, this is done very rarely by the algorithm. In particular, for T total samples drawn by MELK, the design is computed $O(\log_2(T))$ times leading to an overall computational cost of $O(\text{poly}(|\mathcal{X}|)\log_2(T))$ for computing the design. By contrast, any algorithm that computes an acquisition function at every sample suffers computational complexity $\Omega(T)$ for the design. Furthermore, for Gaussian process approaches, the added cost of computing posterior means and variances leads to an overall computational cost of either $\Omega(\text{poly}(|\mathcal{X}|)T)$ or $\Omega(|\mathcal{X}|\text{poly}(T))$ depending on implementation for computing acquisition functions. We focus on the complexity of computing the design and acquisition functions as this is frequently the core computational bottleneck of algorithms for level set estimation and the complexity of drawing samples is usually negligible by comparison. Hence, when many samples are drawn, MELK can be significantly more efficient than past approaches.

Algorithm 3.1. MELK: **M**isspecified **E**xplicit **L**evel set via **K**ernelization

Input: Arms $\mathcal{X}$, ϕ, $\sigma \geq 0$, $\delta > 0$, $\gamma \geq 0$, threshold α, tolerance $\widetilde{\beta}$

$t \leftarrow 1$, $G_1 \to \emptyset$, $B_1 \leftarrow \emptyset$, $\mathcal{A}_1 \leftarrow \mathcal{X}$

while $|G_t \cup B_t| < |\mathcal{X}|$ and $t \leq \lceil \log_2(4/\widetilde{\beta}) \rceil$ **do**

  $\delta_t \leftarrow \delta / 2t^2$

  Let $\lambda_t \in \triangle_{\mathcal{X}}$ minimize $g(\lambda; \mathcal{A}_t; \gamma)$ where

$$g(\lambda; \mathcal{V}; \gamma) := \max_{x \in \mathcal{V}} \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}$$

  $q_t \leftarrow 16 \cdot 2^{2t} g(\lambda_t; \mathcal{A}_t; \gamma)(B^2 + \sigma^2) \log(2t^2 |\mathcal{X}|^2 / \delta)$

  Set $N_t \leftarrow \lceil \max\{q_t, 2\log(|\mathcal{X}|/\delta)\} \rceil$ and sample $x_1, \cdots, x_{N_t}$ observing noisy function values $y_1, \cdots, y_{N_t}$ according to λ_t.

  $\widehat{\theta}_t \leftarrow$ RIPS$(\mathcal{A}_t, \{A^{(\gamma)}(\lambda_t)^{-1} \phi(x_i) y_i\}_{i=1}^{N_t})$, Alg B.1 in Appendix B.2

  for $x \in \mathcal{A}_t$ **do**

    if $\widehat{\theta}_t^\top \phi(x) < \alpha - 2 \cdot 2^{-t}$ **then**

      $B_{t+1} \leftarrow B_t \cup \{x\}$

      $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_t \backslash \{x\}$

    else if $\widehat{\theta}_t^\top \phi(x) > \alpha + 2 \cdot 2^{-t}$ **then**

      $G_{t+1} \leftarrow G_t \cup \{x\}$

      $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_t \backslash \{x\}$

  $t \leftarrow t + 1$

return $\widehat{\mathcal{R}} := \mathcal{X} \setminus B_t$

3.3.2 Optimal Sample Complexity for Explicit Level Set Estimation

Next we state MELK's complexity, deferring constants and doubly logarithmic factors to the appendix for readability.

Theorem 7. *Fix $\delta > 0$, threshold $\alpha > 0$, tolerance $\widetilde{\beta}$, and regularization $\gamma \geq 0$. Define $\Delta_{\min}(\alpha) := \min_{x \in \mathcal{X}} |\phi(x)^T \theta_* - \alpha|$. Define also*

$$\bar{\beta}(\alpha) = \min \left\{ \beta > 0 : 4(\sqrt{\gamma}\|\theta_*\| + h)\Big(2 + \sqrt{\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \in \mathcal{X} : |\phi(x)^\top \theta_* - \alpha| \leq \beta} \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}\Big) \leq \beta \right\}.$$

With probability at least $1 - \delta$, MELK returns a set $\widehat{\mathcal{R}}$ at time T_δ such that

$$\widehat{\mathcal{R}} \supseteq \{x \in \mathcal{X} : f(x) \geq \alpha + \bar{\beta}(\alpha)\} \text{ and } \widehat{\mathcal{R}} \subseteq \{x \in \mathcal{X} : f(x) \geq \alpha - \widetilde{\beta} - \bar{\beta}(\alpha)\}$$

and for any α, $\widetilde{\beta}$ such that $\max(\Delta_{\min}(\alpha), \widetilde{\beta}) \geq \bar{\beta}(\alpha)$

$$T_\delta \leq (B^2 + \sigma^2) \min_{\lambda \in \widetilde{X}} \max_{x \in \mathcal{X}} \frac{\|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{\max\{(\phi(x)^T \theta_* - \alpha)^2, \tilde{\beta}^2\}} \log((\Delta_{\min}(\alpha) \vee \widetilde{\beta})^{-1}) \log\left(|\mathcal{X}| \delta^{-1}\right).$$

We now contextualize the result of our theorem. In the well-specified linear setting with $\phi(x) = x$, $h = 0$, $\widetilde{\beta} = 0$, and $\gamma = 0$ MELK will terminate and return G_α in a time

$$T_\delta \lesssim (B^2 + \sigma^2) \min_{\lambda \in \widetilde{X}} \max_{x \in \mathcal{X}} \frac{\|x\|^2_{A(\lambda)^{-1}}}{(x^T \theta_* - \alpha)^2} \log(\Delta_{\min}^{-1}) \log\left(\frac{|\mathcal{X}|}{\delta}\right)$$

samples which nearly matches the rate suggested by the linear lower bound in Theorem 5. The added factor of $\log(|\mathcal{X}|)$ stems from a union bound, while the dependence on $\log(\Delta_{\min}^{-1})$ is an additional overhead incurred as MELK builds up an estimate of the optimal sample allocation over rounds. We visualize this estimation process in Figure 3.1 in the experiments.

In the more general misspecified setting when $h > 0$, we cannot expect to return G_α exactly and $\overline{\beta}(\alpha)$ characterizes the limit of how well one can estimate $f(x)$. Hence, x's with gaps smaller than $\bar{\beta}(\alpha)$ cannot reliably be detected by MELK. To better understand this quantity, note that for any $\gamma' \in \mathbb{R}$ if we run MELK with $\gamma = \gamma'/T$, Lemma 2 (also Lemma 2 of Camilleri et al. (2021a)) can be used to show that $\bar{\beta}(\alpha) \lesssim (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{\Gamma_T}$ where $\Gamma_T := \sup_{\lambda \in \widetilde{X}} \log\det(TA^{(0)}(\lambda) + \gamma' I)$ is the *maximum information gain* as defined by Srinivas et al. (2009); Gotovos (2013); Bogunovic et al. (2016). Additionally, it can be shown that $\Gamma_T \leq d_{eff}$, where d_{eff} is the *effective dimension* of $\phi(x_1), \ldots, \phi(x_n) \in \mathcal{H}$ as defined in Alaoui and Mahoney (2014); Derezinski et al. (2020). In particular, to ensure that MELK correctly identifies all points that are at least some gap $\Delta > h$ away from the threshold, then we can choose γ so that $\Delta > (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{\Gamma_T}$. In practice we find that $\gamma = 1/T$ works well. Finally, the user may additionally set a tolerance $\widetilde{\beta} > 0$. In this case, we err on the side of potentially returning extra arms that are not in G_α and show that the returned set $\widehat{\mathcal{R}}$ contains all x such that $f(x) > \alpha + \overline{\beta}(\alpha)$ and none such that $f(x) < \alpha - \widetilde{\beta} - \overline{\beta}(\alpha)$. If however, a more selective criteria is desired, the following remark characterizes the output if $\widehat{G}_t$ is returned instead.

Remark. If MELK instead returns $\widehat{\mathcal{R}} = G_t$ then with probability at least $1 - \delta$ $\widehat{\mathcal{R}} \supseteq \{x \in \mathcal{X} : f(x) \geq \alpha + \tilde{\beta} + \bar{\beta}(\alpha)\}$ and and $\widehat{\mathcal{R}} \subseteq \{x \in \mathcal{X} : f(x) \geq \alpha - \bar{\beta}(\alpha)\}$.

Contrast with Existing Approaches. The experimental design based sampling approach is a departure from past work on level set estimation. As opposed to constructing an acquisition function and then bounding the sample complexity of the resulting algorithm as past works have done, we instead begin with an oracle sampling scheme that arises from a lower bound and attempt to design a practical sampling scheme that matches it as more data is collected. In what follows, we compare the guarantees of MELK to the prior art such as Gotovos (2013); Shekhar and Javidi (2019); Bogunovic et al. (2016). As a technical point, we note that these past results are specialized to the Gaussian process setting where a prior on f is known. By contrast, our work presented in this chapter makes no assumption of a prior distribution. Bogunovic (2019); Vakili et al. (2021a) achieve similar guarantees for the frequentist setting. Ignoring these technicalities, our results are tighter than what were previously known.

The past state of the art sample complexities all guarantee that algorithms terminate at the smallest time T satisfying $T \gtrsim \Gamma_T \Delta_{\min}(\alpha)^{-2}$ up to log factors (cf. Thm 1 of (Gotovos, 2013), Cor. 3.1 of (Bogunovic et al., 2016), Thm 1 of (Shekhar and Javidi, 2019), etc.). If we run MELK with $\gamma = \gamma'/T$ then

$$\min_{\lambda \in \triangle_{\mathcal{X}}} \max_x \frac{\|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{(\phi(x)^T\theta_* - \alpha)^2} \leq \min_{\lambda \in \triangle_{\mathcal{X}}} \frac{\max_x \|\phi(x)\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\min_x(\phi(x)^T\theta_* - \alpha)^2} \leq 3\Gamma_T \Delta_{\min}(\alpha)^{-2}$$

where the final inequality follows from Lemma 2 and the definition of $\Delta_{\min}(\alpha)$.

Remark. Combining the above analysis with the result of Theorem 7 highlights that MELK likewise terminates at or before a time T satisfying $T \gtrsim \Gamma_T \Delta_{\min}(\alpha)^{-2}$, though it may stop long before this as the above bound employing Γ_T is only tight in the pathological case when $|\phi(x)^T\theta_* - \alpha| = \Delta_{\min}(\alpha)\ \forall x \in \mathcal{X}$.

Remark. The lower bounds of Scarlett et al. (2017); Cai and Scarlett (2021) show that a dependence $\Omega(\sqrt{\Gamma_T})$ is necessary in the worst case for functions living in an RKHS. Hence, MELK is instance optimal in the linear regime by Theorem 5 and at least minimax optimal in general.

3.4 Implicit Level Set Estimation

In the *implicit* level-set problem, for an $\epsilon \geq 0$ we seek to identify the set $G_\epsilon = \{x : f(x) > (1-\epsilon)f(x_*)\}$. Note that unlike the explicit setting where the threshold α was a given input to the algorithm, now the equivalent notion of a threshold value α is equal to $(1-\epsilon)f(x_*)$, an unknown quantity since it relies on knowledge of the unknown function f. A naive strategy would be to attempt estimate $(1-\epsilon)f(x_*)$ directly and then apply explicit level-set estimation techniques using this estimated threshold value. Indeed, this is precisely the strategy of past works (Mason et al., 2020; Gotovos, 2013). Perhaps surprisingly however, it turns out that estimating the threshold is unnecessary and potentially wasteful. Towards developing lower bound to guide an experimental design, we begin with a simple but powerful observation.

Lemma 4. $x \in G_\epsilon \iff \forall x' \in \mathcal{X} : f(x) \geq (1-\epsilon)f(x')$. *Conversely,* $x \in G_\epsilon^c \iff \exists x' : f(x) < (1-\epsilon)f(x')$.

Proof.

$$x \in G_\epsilon \iff \nexists x' : (1-\epsilon)f(x') > f(x) \iff \forall x' : (1-\epsilon)f(x') \leq f(x)$$

where the second equivalence holds by definition since x_* maximizes $(1-\epsilon)f(x')$ and we have that $f(x) > (1-\epsilon)f(x_*)$ for any $x \in G_\epsilon$. The statement for $x \in G_\epsilon^c$ holds via the negation □

The following corollary specializes the previous lemma to the well specified case.

Corollary 1. *In the well specified setting where* $h = 0$*,*

$$x \in G_\epsilon \iff \forall x' \in \mathcal{X} : \theta_*^\top(\phi(x) - (1-\epsilon)\phi(x')) \geq 0$$

and conversely,

$$x \in G_\epsilon^c \iff \exists x' : \theta_*^\top(\phi(x) - (1-\epsilon)\phi(x')) < 0.$$

This lemma highlights that to determine if $x \in G_\epsilon$, one need only check if

$$\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x')) > 0 \text{ for all } x' \in \mathcal{X}.$$

In particular, this does not require any estimate of the threshold $(1-\epsilon)f(x^*)$. Instead, it is only necessary to maintain estimates of ordered pairs of points (x, x') without searching for x_* directly. Next, to guide our algorithm design we look to an information-theoretic lower bound.

Theorem 8. *In the well-specified linear setting when* $\phi(x) = x$ *and* $f(x) = \theta_*^\top x$*, for any* $\delta > 0$*, any algorithm that returns the set* G_ϵ *with probability at least* $1-\delta$ *must satisfy*

$$\frac{\mathbb{E}[T_\delta]}{\log(1/2.4\delta)} \geq 2 \min_{\lambda \in \triangle_\mathcal{X}} \max \left\{ \max_{z \in G_\epsilon} \max_{x' \in \mathcal{X}} \frac{\|x-(1-\epsilon)x'\|^2_{A(\lambda)^{-1}}}{(\theta_*^\top(x-(1-\epsilon)x'))^2}, \max_{x \in G_\epsilon^c} \min_{x' \in \mathcal{X}} \frac{\|x-(1-\epsilon)x'\|^2_{A(\lambda)^{-1}}}{(\theta_*^\top(x-(1-\epsilon)x'))^2} \right\}$$

where T_δ *denotes the random stopping time.*

Notably, the directions $\phi(x) - (1-\epsilon)\phi(x')$ naturally arise in the lower bound. This suggests an optimal sampling distribution λ^* that achieves the minimum of the inequality in 8. As was the case in explicit level set estimation, this sampling distribution also depends on the unknown θ_*.

3.4.1 Algorithm

Motivated by the lower bound, we propose Algorithm 3.2 called MILK (**M**isspecified **I**mplicit **L**evel set via **K**ernelization) which proceeds in phases where we attempt to progressively match the optimal distribution from the lower bound as was done by MELK for the explicit setting. The key difference, however is that MILK instead computes a design to optimally estimate $\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x'))$ rather than $\theta_*^\top\phi(x)$ as in MELK. Given active set $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ of *pairs* of arms define,

$$\mathcal{Y}^\epsilon(\mathcal{A}) := \{\phi(x) - (1-\epsilon)\phi(x') : (x, x') \in \mathcal{A}\}.$$

The active set in round 1 is initialized as $\mathcal{A}_1 = \mathcal{X} \times \mathcal{X}$. MILK keeps track of sets $\widehat{G}_t \subset \mathcal{X}$ and $\widehat{B}_t \subset \mathcal{X}$ of arms it believes to be in G_ϵ and G_ϵ^c and makes use of the RIPS procedure to robustly estimate means. As the algorithm proceeds, in each round t an optimal design is computed over remaining difference vectors in $\mathcal{Y}^\epsilon(\mathcal{A}_t)$ and the number of samples N_t is sufficient to ensure that $|(\theta_* - \widehat{\theta})^\top(\phi(x) - (1-\epsilon)\phi(x'))| \le 2^{-t+1}$. Then for every arm that has not been added to $\widehat{G}_t$ or $\widehat{B}_t$, MILK does the following:

$$\text{if } \exists x' : \widehat{\theta}^\top((\phi(x) - (1-\epsilon)\phi(x')) < 2^{-t}$$

then x is added to $\widehat{B}_t$. In our proof, we show this condition occurs if and only if there exists a x' such that $\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x')) < 0$. If this occurs, all pairs of the form (x, x') or (x', x), $x' \in \mathcal{X}$ are removed from $\mathcal{A}_t$[2]. Semantically, if MILK can ensure that x is not in G_ϵ, then x is never sampled again. Otherwise, for any x' if $\widehat{\theta}^\top(\phi(x) - (1-\epsilon)\phi(x') > 2^{-t}$, the single pair (x, x') is removed from $\mathcal{A}_t$. An arm x is only ever added to $\widehat{G}_t$ if $\{(x, x'), x' \in \mathcal{X}\} \cap \mathcal{A}_t = \emptyset$ which occurs when

$$\forall x' : \exists t' \text{ such that } \widehat{\theta}_{t'}^\top((\phi(x) - (1-\epsilon)\phi(x')) > 2^{-t'}.$$

In our proof, we show that this occurs if and only if $\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x')) > 0$ for all $x' \in \mathcal{X}$ which is both necessary and sufficient by Lemma 4. Note that even if x has been added to $\widehat{G}_t$ implying that all pairs (x, x') have been removed from $\mathcal{A}_t$, x may be present in other pairs (x', x) which can be necessary to determine if $x' \in G_\epsilon$. Finally, the algorithm terminates when either every arm has been added to either $\widehat{G}_t$ or $\widehat{B}_t$ or it has reached a round $t \gtrsim \log_2(1/\widetilde{\beta})$ when the desired tolerance $\widetilde{\beta}$ is achieved.

3.4.2 Theoretical Guarantees

Next we state MILK's complexity, again deferring constants and doubly logarithmic factors to the appendix for readability.

Theorem 9. *Fix $\delta > 0$, $\epsilon > 0$, tolerance $\widetilde{\beta}$, and regularization $\gamma > 0$. Define $\Delta_{\min}(\epsilon) = \min_x |\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x^*))|$. Define also*

$$\bar{\beta}(\epsilon) = \min_{\beta > 0} \left\{ 4(\sqrt{\gamma}\|\theta_*\| + h)\left(2 + \sqrt{\min_{\lambda \in \widetilde{X}} \nu(\lambda, \beta)}\right) \le \beta \right\},$$
$$\nu(\lambda, \beta) := \max_{\substack{(x,x') \in \mathcal{X}\times\mathcal{X} \\ |\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x'))| \le \beta}} \|\theta_*^\top(\phi(x) - (1-\epsilon)\phi(x'))\|^2_{A^{(\gamma)}(\lambda)^{-1}}.$$

[2]We assume that pairs are ordered, i.e. $(x, x') \neq (x', x)$ for $x \neq x'$.

Algorithm 3.2. MILK: **M**isspecified **I**mplicit **L**evel set via **K**ernelization

Input: Arms $\mathcal{X}$, ϕ, $\delta > 0$, $\epsilon > 0$, $\gamma \geq 0$, tolerance $\widetilde{\beta}$

$t \leftarrow 1$, $\widehat{G}_1 \to \emptyset$, $\widehat{B}_1 \leftarrow \emptyset$, $\mathcal{A}_1 \leftarrow \{(x, x'), x, x' \in \mathcal{X}\}$

while $|\widehat{G}_t \cup \widehat{B}_t| < |\mathcal{X}|$ and $t \leq \lceil \log_2(4/\widetilde{\beta}) \rceil$ **do**

  $\delta_t \leftarrow \delta / 2t^2$

  Let $\lambda_t \in \triangle_{\mathcal{X}}$ minimize $g(\lambda; \mathcal{A}_t; \gamma)$ where

$$g(\lambda, \mathcal{V}; \gamma) := \max_{(x,x') \in \mathcal{V}} \|\phi(x) - (1-\epsilon)\phi(x')\|^2_{A^{(\gamma)}(\lambda)^{-1}}$$

  $q_t \leftarrow 16 \cdot 2^{2t} g(\lambda_t; \mathcal{A}_t; \gamma)(B^2 + \sigma^2) \log(2t^2 |\mathcal{X}|^2 / \delta)$

  Set $N_t \leftarrow \lceil \max\{q_t, 2\log(|\mathcal{X}|/\delta)\} \rceil$ and sample $x_1, \cdots, x_{N_t}$ observing noisy function values $y_1, \cdots, y_{N_t}$ according to λ_t.

  $\widehat{\theta}_t \leftarrow \text{RIPS}(\mathcal{Y}^\epsilon(\mathcal{A}_t), \{A^{(\gamma)}(\lambda_t)^{-1} \phi(x_i) y_i\}_{i=1}^{N_t})$

  for $(x, x') \in \mathcal{A}_t$ **do**

    if $\widehat{\theta}_t^\top (\phi(x) - (1-\epsilon)\phi(x')) < -2 \cdot 2^{-t}$ **then**

      $\widehat{B}_{t+1} \leftarrow x$

      x-pairs $\leftarrow \{(x, x') \text{ and } (x', x) | x' \in \mathcal{X}\}$

      $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_t \setminus x$-pairs

    if $\widehat{\theta}_t^\top (\phi(x) - (1-\epsilon)\phi(x')) > 2 \cdot 2^{-t}$ **then**

      $\mathcal{A}_{t+1} \leftarrow \mathcal{A}_t \setminus \{(x, x')\}$

      if $\{(x, x') | x' \in \mathcal{X}\} \cap \mathcal{A}_t = \emptyset$ **then**

        $\widehat{G}_{t+1} \leftarrow \widehat{G}_t \cup \{x\}$

  $t \leftarrow t + 1$

return $\widehat{\mathcal{R}} := \mathcal{X} \setminus \widehat{B}_t$

With probability $1 - \delta$, MILK *returns a set* $\widehat{\mathcal{R}}$ *at a time* T_δ *such that*

$$\widehat{\mathcal{R}} \supseteq \{x \in \mathcal{X} : f(x) \geq (1-\epsilon) f(x_*) + \bar{\beta}(\epsilon)\} \text{ and}$$
$$\widehat{\mathcal{R}} \subseteq \{x \in \mathcal{X} : f(x) \geq (1-\epsilon) f(x_*) - \widetilde{\beta} - \bar{\beta}(\epsilon)\}$$

and for any ϵ, $\widetilde{\beta}$ *such that* $\max(\Delta_{\min}(\epsilon), \widetilde{\beta}) \geq \bar{\beta}(\epsilon)$

$$T_\delta \leq (B^2 + \sigma^2) H^{MILK}(\theta_*) \log_2((\Delta_{\min}(\epsilon) \vee \widetilde{\beta})^{-1}) \log\left(\frac{|\mathcal{X}|}{\delta}\right)$$

for $H^{MILK}(\theta_*) = \min_{\lambda \in \widetilde{X}} \left\{ H_\lambda^{MILK\text{-}G_\epsilon}(\theta_*) \vee H_\lambda^{MILK\text{-}G_\epsilon^c}(\theta_*) \right\}$, *where*

$$H_\lambda{}^{MILK\text{-}G_\epsilon}(\theta_*) := \max_{x \in G_\epsilon} \max_{x' \in \mathcal{X}} \frac{\|\phi(x) - (1-\epsilon)\phi(x')\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{\max\{((\phi(x) - (1-\epsilon)\phi(x'))^\top \theta_*)^2, \widetilde{\beta}^2\}},$$

$$\text{and } H_\lambda^{MILK\text{-}G_\epsilon^c}(\theta_*) := \max_{x \in G_\epsilon^c} \max_{x'} \frac{\|\phi(x) - (1-\epsilon)\phi(x')\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{\max\{((\phi(x) - (1-\epsilon)\phi(x_*))^\top \theta_*)^2, \widetilde{\beta}^2\}}.$$

The statement of Theorem 9 for MILK is similar that of 7 for MELK. In the well specified case when $\widetilde{\beta} = 0$, MILK returns G_ϵ exactly at a time T_δ that satisfies

$$T_\delta \lesssim (B^2 + \sigma^2) H^{\texttt{MILK}}(\theta_*) \log_2(\Delta_{\min}(\epsilon)) \log\left(|\mathcal{X}|\delta^{-1}\right)$$

In this case, however, $H^{\texttt{MILK}}(\theta_*)$ is a maximum of two different complexity terms. $H_\lambda^{\texttt{MILK}-G_\epsilon}$ represents the complexity of identifying all $x \in G_\epsilon$. Similarly, $H_\lambda^{\texttt{MILK}-G_\epsilon^c}$ represents the complexity of identifying all $x \in G_\epsilon^c$. Similar to the explicit setting, in the misspecified case when $h > 0$, $\overline{\beta}(\epsilon)$ similarly represents the limit of how well we can estimate $f(x)$ for any $x \in \mathcal{X}$ and $\widetilde{\beta}$ allows for an additional tolerance such that MILK detects all x for which $f(x) > (1-\epsilon)f(x_*) + \overline{\beta}(\epsilon)$ and none worse than $f(x) < (1-\epsilon)f(x_*) - \overline{\beta}(\epsilon) - \widetilde{\beta}$. The following remark addresses the setting where MILK returns $\widehat{G}_t$ instead.

Remark: If the algorithm instead returns $\widehat{\mathcal{R}} = \widehat{G}_t$, then with probability at least $1 - \delta$

$$\widehat{\mathcal{R}} \supseteq \{x \in \mathcal{X} : f(x) \geq (1-\epsilon)f(x_*) + \widetilde{\beta} + \bar{\beta}(\epsilon)\} \text{ and}$$
$$\widehat{\mathcal{R}} \subseteq \{x \in \mathcal{X} : f(x) \geq (1-\epsilon)f(x_*) - \bar{\beta}(\epsilon)\}.$$

Comparison with the Lower Bound

The complexity term $H^{\texttt{MILK}}(\theta_*)$ naturally breaks into two terms. $H^{\texttt{MILK}-G_\epsilon}(\theta_*)$ represents the complexity of finding arms in G_ϵ and it matches a corresponding term in the lower bound. $H^{\texttt{MILK}-G_\epsilon^c}(\theta_*)$ represents the complexity of removing arms in G_ϵ^c but is slightly different than the term in the lower bound. As a consequence of Theorem 4.1 of Mason et al. (2020) however, one can show the term given in the lower bound for $x \in G_\epsilon^c$ is not achievable except asymptotically as $\delta \to 0$ in general. Instead, the problem of implicit level set estimation reduces to the problem of all ϵ-good arm identification in multi-armed bandits studied by Mason et al. (2020) when $\phi(x) = x$, $h = 0$, and $x_i = e_i$. We show in the appendix that MILK's sample complexity matches the optimal finite time rate up to logarithmic factors as shown in Mason et al. (2020).

Contrast with Existing Results

As was shown in the explicit setting, we can show that the sample complexity bound in Theorem 9 improves on the current state of the art. Take $\gamma = \gamma'/T$ for any $\gamma' \in \mathbb{R}$. Then we may bound $H^{\texttt{MILK}-G_\epsilon}(\theta_*)$ as

$$\begin{aligned}
&\min_{\lambda \in \widetilde{X}} \max_{x,x'} \left\{ \frac{\|\phi(x) - (1-\epsilon)\phi(x')\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{((\phi(x) - (1-\epsilon)\phi(x'))^\top \theta_*)^2} \right\} \\
&\quad \overset{(a)}{\leq} 4 \min_{\lambda \in \widetilde{X}} \max_{x,x'} \left\{ \frac{(1-\epsilon)^2 \|\phi(x) - \phi(x')\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{((\phi(x) - (1-\epsilon)\phi(x'))^\top \theta_*)^2} \ \vee \ \frac{\epsilon^2 \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{((\phi(x) - (1-\epsilon)\phi(x'))^\top \theta_*)^2} \right\} \\
&\quad \overset{(b)}{\leq} 4 \frac{(1+\epsilon)^2}{\Delta_{\min}(\epsilon)^2} \min_{\lambda \in \widetilde{X}} \max_{x,x'} \left\{ \|\phi(x') - \phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}} \vee \|\phi(x)\|^2_{(A^{(\gamma)}(\lambda))^{-1}} \right\} \\
&\quad \leq \frac{8(1+\epsilon)^2}{\Delta_{\min}(\epsilon)^2} \min_{\lambda \in \widetilde{X}} \max_{x} \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}} \\
&\quad \overset{(c)}{\leq} \frac{12(1+\epsilon)^2}{\Delta_{\min}(\epsilon)^2} \Gamma_T
\end{aligned}$$

where (a) follows by the triangle inequality, (b) by definition of $\Delta_{\min}(\epsilon)$ and (c) follows by Lemma 2. A similar computation follows for $H^{\texttt{MILK}-G_\epsilon^c}(\theta_*)$ Hence, MILK is at most $O(\Gamma_T \Delta_{\min}^{-2})$ though it can be much tighter as inequality (b) is tight only in the worst case when all gaps are equal. In particular, the result of Theorem 9 is tighter than Theorem 2 of Gotovos (2013).

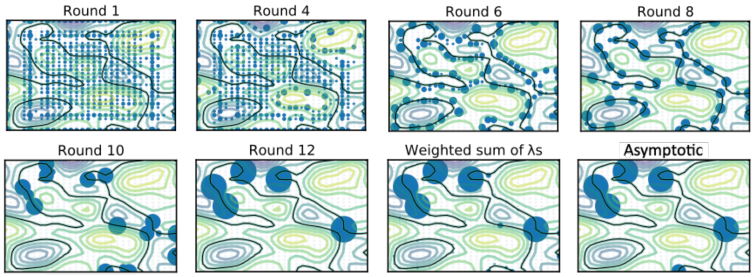
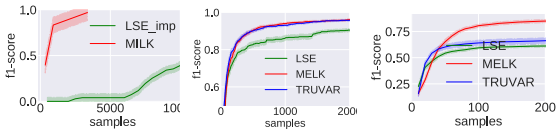


Figure 3.1: Allocations across rounds for a function $f(x, y)$ with a threshold of $\alpha = 0$ shown in black.



(a) Implicit **(b)** GP, $\ell = 0.05$ **(c)** Cosine

Figure 3.2: Performance of `MELK` and `MILK` versus Gaussian process baseline.

3.5 Experiments

In this section, we compare our algorithms to existing baselines in the literature. Additional details of these methods and our experiments are in the Appendix.

Warm-Up: Optimal Sampling. In Figure 3.1 we illustrate the sampling behavior of `MELK`. We let $\mathcal{X} = \{(\frac{i}{30}, \frac{j}{30})\}_{i,j=1}^{30}$ and considered the squared exponential kernel $k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\|\boldsymbol{x} - \boldsymbol{x}'\|^2/2\ell^2)$ with parameter $\ell = 0.1$. We also chose $\theta_* \sim \mathcal{N}(0, I_{900})$ and show a contour plot of $f(\boldsymbol{x}) = \theta_*^\top \phi(\boldsymbol{x})$. The black curve represents the boundary of the $\alpha = 0$ level set. We plot the sample allocations as the algorithm progresses (taking $\gamma = 0$). The initial distribution is mostly uniform with several sampling modes. In later rounds, the points nearest to the boundary of the level set, given by the black curve are sampled, and eventually, only the points with the smallest gaps (the most difficult regions) receive samples. As the number of samples in round t is proportional to 2^{2t}, we compute the sum of the designs weighted by the 2^{2t} to show the overall sampling design. Additionally, we plot the asymptotic allocation suggested by Theorem 5, namely $\lambda_* = \arg\min_\lambda \max_{x \in X} \|\phi(\boldsymbol{x})\|^2_{A^{(\gamma)}(\lambda)^{-1}}/(\theta_*^\top \phi(\boldsymbol{x}) - \alpha)^2$. In particular, the weighted sum of the designs taken by `MELK` is nearly identical to λ_*.

Gaussian Process Level Set Estimation. For our main empirical evaluation, we focused on the Gaussian Process setting for the explicit level set problem. In the explicit level-set case we compare to LSE (Gotovos, 2013) and TruVar (Bogunovic et al., 2016). We drew a function $f : [0, 1] \to \mathbb{R}$ from the Gaussian process $\mathcal{N}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ where the kernel is a squared exponential kernel with parameter $\ell = .05$ and $[0, 1]$ was uniformly discretized into 200 points. We assumed that the noise variance was $\sigma^2 = 1$ (high noise) and the threshold was chosen so that 10% of the function values were above it. In this setting, we implement a batched version of `MELK` that draws a fixed batch size of samples each round (namely 10) and then recomputes the design. This reflects the practical constraint that experimenters may wish to collect a fixed number of samples at a time rather than a potentially growing amount. To provide a fair comparison to the GP-based methods, we computed a posterior distribution on f in each round. For each point we replaced our theoretically justified confidence intervals in the RKHS setting with confidence intervals arising from the posterior, namely $\hat{\mu}_t(\boldsymbol{x}) \pm \beta^{1/2}\hat{\sigma}_t(\boldsymbol{x})$ where $\hat{\mu}_t, \hat{\sigma}_t$ are the posterior mean and standard deviations respectively. As in past works, we take $\beta^{1/2} = 3$ as theoretically justified choices of β (eg. Theorem 1 of (Srinivas et al., 2009)) tend to be overly conservative. We also took γ dropping like $1/i$ on the i-th round we computed the design. We ran 25 repetitions drawing a new choice of f each run. Figure 3.2b shows the average F1 score

of the set of points each algorithm declares to be in G_α respectively with bars denoting 1 standard error. Our algorithm performs very similarly to `TruVar` - an algorithm whose acquisition function samples in a way to reduce the average variance, unlike our method which tries to reduce the maximum variance.

Our second comparison is in Figure 3.2c: we took $f(x) = \cos(8\pi \boldsymbol{x})$, $\ell = .1$, $\sigma = .2$ (low noise regime) and chose the threshold so that 30% of points were above it. We then considered 700 points uniformly in $[0, 1]$. In the appendix, we vary the underlying parameters of ℓ, σ^2 to demonstrate the performance of these algorithms in different regimes.

Linear Implicit Case. We additionally compare against `LSE-imp` in the linear setting where $\phi(\boldsymbol{x}) = \boldsymbol{x}$ on a benchmark example from the linear bandits literature designed to test the effectiveness of adaptive sampling algorithms (Soare et al., 2014). For $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^d$, we take $\boldsymbol{x}_1 = \boldsymbol{x}_* = \theta_* = e_1$ and $\boldsymbol{x}_2 = e_2$. The remaining $\boldsymbol{x}_3, \cdots, \boldsymbol{x}_n$ are set so that their first two coordinates are $\cos(\pi/4(1 + \xi))e_1$ and $\sin(\pi/4(1 + \xi))e_2$ for $\xi \sim$ Unif$(-.2, .2)$. We set the threshold $\alpha = 0.5$, $n = 100$, and $d = 25$. Though it is far below α, sampling arm $\boldsymbol{x}_2$ provides the most information about which arms exceed the threshold. In this setting, we ran both algorithms with the exact confidence intervals as specified by their respective theoretical guarantees leading to large sample complexities, and we include further details in the appendix. Indeed, we see in 3.2a that `MILK` outperforms `LSE-imp`.

3.6 Conclusion

In this chapter, we provide the first instance optimal algorithms for explicit and implicit level set estimation and provide theoretical and empirical justification for our algorithms.

Chapter 4

Selective Sampling for Online Best-arm Identification

4.1 Introduction

In this chapter we consider *selective sampling for online best-arm identification*. In this setting, at every time step $t = 1, 2, \ldots$, Nature reveals a potential measurement $x_t \in \mathcal{X} \subset \mathbb{R}^d$ to the learner. The learner can choose to either *query* x_t ($\xi_t = 1$) or abstain ($\xi_t = 0$) and immediately move on to the next time. If the learner chooses to take a query ($\xi_t = 1$), then Nature reveals a noisy linear measurement of an unknown $\theta_* \in \mathbb{R}^d$, i.e. $y_t = \langle x_t, \theta_* \rangle + \epsilon_t$ where ϵ_t is mean zero sub-Gaussian noise. Before the start of the game, the learner has knowledge of a set $\mathcal{Z} \subset \mathbb{R}^d$. The objective of the learner is to identify $z_* := \arg\max_{z \in \mathcal{Z}} \langle z, \theta_* \rangle$ with probability at least $1 - \delta$ at a learner specified stopping time $\mathcal{U}$. It is desirable to minimize both the stopping time $\mathcal{U}$ which counts the total number of unlabeled or labeled queries and the number of labeled queries requested $\mathcal{L} := \sum_{t=1}^{\mathcal{U}} \mathbf{1}\{\xi_t = 1\}$. In this setting, at each time t the learner must make the decision of whether to accept the available measurement x_t, or abstain and wait for an even more informative measurement. While abstention may result in a smaller total labeled sample complexity $\mathcal{L}$, the stopping time $\mathcal{U}$ may be very large. This chapter characterizes the set of feasible pairs $(\mathcal{U}, \mathcal{L})$ that are necessary and sufficient to identify z_* with probability at least $1 - \delta$ when x_t are drawn IID at each time t from a distribution ν. Moreover, we propose an algorithm that nearly obtains the minimal information theoretic label sample complexity $\mathcal{L}$ for any desired unlabeled sample complexity $\mathcal{U}$.

While characterizing the sample complexity of selective sampling for online best arm identification is the primary theoretical goal of the work presented in this chapter, the study was initially motivated by fundamental questions about how to optimally trade-off the value of information versus time. Even for this idealized linear setting, it is far from obvious a priori what an optimal decision rule ξ_t looks like and if it can even be succinctly described, or if it is simply the solution to an opaque optimization problem. Remarkably, we show that for every feasible, optimal operating pair $(\mathcal{U}, \mathcal{L})$ there exists a matrix $A \in \mathbb{R}^{d \times d}$ such that the optimal decision rule takes on the form $\xi_t = 1\{x^\top A x \geq 1\}$ when $x_t \sim \nu$ iid. The fact that for any smooth distribution ν the decision rule is a hard decision equivalent to x_t falling outside a fixed ellipse or not, and not a stochastic rule that varies complementarily with the density of ν over space is perhaps unexpected.

To motivate the problem description, suppose on each day $t = 1, 2, \ldots$ a food blogger posts the *Cocktail of the Day* with a recipe described by a feature vector $x_t \in \mathbb{R}^d$. You have the ingredients (and skills) to make any possible cocktail in the space of all cocktails $\mathcal{Z}$, but you don't know which one you'd like the most, i.e., $z_* := \arg\max_{z \in \mathcal{Z}} \langle z, \theta_* \rangle$, where θ_* captures your preferences over cocktail recipes. You decide to use the

Cocktail of the Day to inform your search. That is, each day you are presented with the cocktail recipe $x_t \in \mathbb{R}^d$, and if you choose to make it ($\xi_t = 1$) you observe your preference for the cocktail y_t with $\mathbb{E}[y_t] = \langle x_t, \theta_* \rangle$. Of course, making cocktails can get costly, so you don't want to make each day's cocktail, but rather you will only make the cocktail if x_t is informative about θ_* (e.g., uses a new combination of ingredients). At the same time, waiting too many days before making the next cocktail of the day may mean that you never get to learn (and hence drink) the cocktail z_* you like best. The setting above is not limited to cocktails, but rather naturally generalizes to discovering the efficacy of drugs and other therapeutics where blood and tissue samples come to the clinic in a stream and the researcher has to choose whether to take a potentially costly measurement.

Our results hold for arbitrary $\theta_* \in \mathbb{R}^d$, sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Z} \subset \mathbb{R}^d$, and measures $\nu \in \triangle_{\mathcal{X}}$[1] for which we assume $x_t \sim \nu$ is drawn IID. The assumption that each x_t is IID allows us to make very strong statements about optimality. To summarize, our contributions are as follows:

- We present fundamental limits on the trade-off between the amount of unlabelled data and labelled data in the form of (the first) information theoretic lower bounds for selective sampling problems that we are aware of. Naturally, they say that there is an absolute minimum amount of unlabelled data that is necessary to solve the problem, but then for any amount of unlabelled data beyond this critical value, the bounds say that the amount of labelled data must exceed some value as a function of the unlabelled data used.

- We propose an algorithm that nearly matches the lower bound at all feasible trade-off points in the sense that given any unlabelled data budget that exceeds the critical threshold, the algorithm takes no more labels than the lower bound suggests. Thus, the upper and lower bounds sketch out a curve of all possible operating points, and the algorithm achieves any point on this curve.

- We characterize the optimal decision rule of whether to take a sample or not, based on any critical point is a simple test: Accept $x_t \in \mathbb{R}^d$ if $x_t^\top A x_t \geq 1$ for some matrix A that depends on the desired operating point and geometry of the task. Geometrically, this is equivalent to x_t falling inside or outside an ellipsoid.

- Our framework is also general enough to capture binary classification, and consequently, we prove results there that improve upon state of the art.

4.1.1 Related Work

Selective Sampling in the Streaming Setting: Online prediction, the setting in which the selective sampling framework was introduced, is a closely related problem to the one studied in our work and enjoys a much more developed literature (Cesa-Bianchi et al., 2009; Dekel et al., 2012; Agarwal, 2013; Chen et al., 2021). In the linear online prediction setting, for $t = 1, 2, \ldots$ Nature reveals $x_t \in \mathbb{R}^d$, the learner predicts $\widehat{y}_t$ and incurs a loss $\ell(\widehat{y}_t, y_t)$, and then the learner decides whether to observe y_t (i.e., $\xi_t = 1$) or not ($\xi_t = 0$), where y_t is a label generated by a composition of a known link function with a linear function of x_t. For example, in the classification setting (Agarwal, 2013; Cesa-Bianchi et al., 2009; Dekel et al., 2012), one setting assumes $y_t \in \{-1, 1\}$ with $\mathbb{E}[y_t | x_t] = \langle x_t, \theta_* \rangle$ for some unknown $\theta_* \in \mathbb{R}^d$, and $\ell(\widehat{y}_t, y_t) = \mathbf{1}\{\widehat{y}_t \neq y_t\}$. In the regression setting (Chen et al., 2021), one observes $y_t \in [-1, 1]$ with $\mathbb{E}[y_t | x_t] = \langle x_t, \theta_* \rangle$ again, and $\ell(\widehat{y}_t, y_t) = (\widehat{y}_t - y_t)^2$. After any amount of time $\mathcal{U}$, the learner is incentivized to minimize both the amount of requested labels $\sum_{t=1}^{\mathcal{U}} \mathbf{1}\{\xi_t = 1\}$ and the cumulative loss $\sum_{t=1}^{\mathcal{U}} \ell(y_t, \widehat{y}_t)$ (or some measure of regret

[1] We denote the set of probability measures over $\mathcal{X}$ as $\triangle_{\mathcal{X}}$.

which compares to predictions using the unknown θ_*). If every label y_t is requested then $\mathcal{L} = \mathcal{U}$ and this is just the classical online learning setting.

These works give a guarantee on the regret and labeled points taken in terms of the hardness of the stream relative to a learner which would see the label at every time. Most do not give the learner the ability to select an operating point that provides a trade-off between the amount of unlabeled versus labeled data taken. Those few works that propose algorithms that do provide this functionality do not provide lower bounds that match their given upper bounds, leaving it unclear whether their algorithm optimally negotiates this trade-off. In contrast, our work fully characterizes the trade-off between the amount of unlabeled and labeled data through an information-theoretic lower bound and a matching upper bound. Specifically, our algorithm includes a tuning parameter, call it τ, that controls the trade-off between the evaluation metric of interest (for us, the quality of the recommended $z \in \mathcal{Z}$), the label complexity $\mathcal{L}$, and the amount of unlabelled data $\mathcal{U}$ that is necessary before the metric of interest can be non-trivial. We prove that each possible setting of τ parametrizes *all* possible trade-offs between unlabeled and labeled data.

Our work is perhaps closest to the streaming setting for agnostic active classification (Dasgupta et al., 2008; Huang et al., 2015) where each x_s is drawn i.i.d. from an underlying distribution ν on $\mathcal{X}$, and indeed our results can be specialized to this setting as we discuss in Section 4.3. These papers also evaluate themselves at a single point on the tradeoff curve, namely the number of samples needed in passive supervised learning to obtain a learner with excess risk at most ϵ. They provide minimax guarantees on the amount of labeled data needed in terms of the disagreement coefficient (Hanneke et al., 2014). In contrast, again, our results characterize the full trade-off between the amount of unlabeled data seen, and the amount of labeled data needed to achieve the target excess risk ϵ. We note that using online-to-batch conversion methods, (Dekel et al., 2012; Agarwal, 2013; Cesa-Bianchi et al., 2009) also provide results on the amount of labeled data needed but they assume a very specific parametric form to their label distribution unlike our setting which is agnostic. Other works have characterized selective sampling for classification in the realizable setting that assumes there exists a classifer among the set under consideration that perfectly labels every y_t (Hanneke and Yang, 2021)–our work addresses the agnostic setting where no such assumption is made. Finally, our results apply under the more general setting of *domain adaptation under covariate shift* where we are observing data drawn from the stream ν, but we will evaluate the excess risk of our resulting classifier on a different stream π (Rai et al., 2010; Saha et al., 2011; Xiao and Guo, 2013).

Best-Arm Identification and Online Experimental Design. Our techniques are based on experimental design methods for best-arm identification in linear bandits, see Chapter 2 or (Soare et al., 2014; Fiez et al., 2019; Camilleri et al., 2021a). In the setting of these works, there exists a pool of examples $\mathcal{X}$ and at each time any $x \in \mathcal{X}$ can be selected with replacement. The goal is to identify the best arm using as few total selections (labels) as possible. Their algorithms are based on arm-elimination. Specifically, they select examples with probability proportional to an approximate G-optimal design with respect to the current remaining arms. Then, during each round after taking measurements, those arms with high probability of being suboptimal will be eliminated. Remarkably, near-optimal sample complexity has been achieved under this setting. While we apply these techniques of arm-elimination and sampling through G-optimal design, the major difference is that we are facing a stream instead of a pool of examples. Finally, (Eghbali et al., 2018) considers a different online experiment design setup where (adversarially chosen) experiments arrive sequentially and a primal-dual algorithm decides whether to choose each, subject to a total budget. (Eghbali et al., 2018) studies the competitive ratio of such algorithms (in the manner of online packing algorithms) for problems such as D-optimal experiment design.

4.2 Selective Sampling for Best Arm Identification

Consider the following game: Given known $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$ and unknown $\theta_* \in \mathbb{R}^d$ at each time $t = 1, 2, \ldots$:

1. Nature reveals $x_t \overset{iid}{\sim} \nu$ with support$(\nu) = \mathcal{X}$
2. Player chooses $Q_t \in \{0, 1\}$. If $Q_t = 1$ then nature reveals y_t with $\mathbb{E}[y_t] = \langle x_t, \theta_* \rangle$
3. Player optionally decides to stop at time t and output some $\widehat{z} \in \mathcal{Z}$

If the player stops at time $\mathcal{U}$ after observing $\mathcal{L} = \sum_{t=1}^{\mathcal{U}} Q_t$ labels, the objective is to identify $z_* = \arg\max_{z \in \mathcal{Z}} \langle z, \theta_* \rangle$ with probability at least $1 - \delta$ while minimizing a trade-off of $\mathcal{U}, \mathcal{L}$.

This chapter studies the relationship between $\mathcal{U}$ and $\mathcal{L}$ in the context of necessary and sufficient conditions to identify z_* with probability at least $1 - \delta$. Clearly $\mathcal{U}$ must be "large enough" for z_* to be identifiable even if all labels are requested (i.e., $\mathcal{L} = \mathcal{U}$). But if $\mathcal{U}$ is very large, the player can start to become more picky with their decision to observe the label or not. Indeed, one can easily imagine scenarios in which it is advantageous for a player to forgo requesting the label of the current example in favor of waiting for a more informative example to arrive later if they wished to minimize $\mathcal{L}$ alone. Intuitively, $\mathcal{L}$ should decrease as $\mathcal{U}$ increases, but how?

Any selective sampling algorithm for the above protocol at time t is defined by 1) a selection rule $P_t : \mathcal{X} \to [0, 1]$ where $Q_t \sim$ Bernoulli$(P_t(x_t))$, 2) a stopping rule $\mathcal{U}$, and 3) a recommendation rule $\widehat{z} \in \mathcal{Z}$. The algorithm's behavior at time t can use all information collected up to time t

Definition 2. *For any $\delta \in (0, 1)$ we say a selective sampling algorithm is δ-PAC for $\nu \in \triangle_{\mathcal{X}}$ if for all $\theta \in \mathbb{R}^d$ the algorithm terminates at time $\mathcal{U}$ which is finite almost surely and outputs $\arg\max_{z \in \mathcal{Z}} \langle z, \theta \rangle$ with probability at least $1 - \delta$.*

4.2.1 Optimal design

Before introducing our own algorithm, let us consider a seemingly optimal procedure. For any $\lambda \in \triangle_{\mathcal{X}} = \{p : \sum_{x \in \mathcal{X}} p_x = 1,\ p_x \geq 0\ \forall x \in \mathcal{X}\}$ define

$$\rho(\lambda) := \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z - z_*\|^2_{\mathbb{E}_{X \sim \lambda}[XX^\top]^{-1}}}{\langle \theta_*, z_* - z \rangle^2}. \tag{4.1}$$

Intuitively, $\rho(\lambda)$ captures the number of labeled examples drawn from distribution λ to identify z_*. Specifically, for any $\tau \geq \rho(\lambda) \log(|\mathcal{Z}|/\delta)$, if $x_1, \ldots, x_\tau \sim \lambda$ and $y_i = \langle x_i, \theta_* \rangle + \epsilon_i$ where ϵ_i is iid 1 sub-Gaussian noise, then there exists an estimator $\widehat{\theta} := \widehat{\theta}(\{(x_i, y_i)\}_{i=1}^\tau)$ such that $\langle \widehat{\theta}, z_* \rangle > \max_{z \in \mathcal{Z} \setminus z_*} \langle \widehat{\theta}, z \rangle$ with probability at least $1 - \delta$ (Fiez et al., 2019). In particular, $\tau \geq \rho(\lambda) \log(|\mathcal{Z}|/\delta)$ samples suffice to guarantee that $\arg\max_{z \in \mathcal{Z}} \langle \widehat{\theta}, z \rangle = \arg\max_{z \in \mathcal{Z}} \langle \theta_*, z \rangle =: z_*$.

Thus, if our τ samples are coming from ν, we would expect any reasonable algorithm to require at least $\rho(\nu) \log(|\mathcal{Z}|/\delta)$ examples and labels. However, since we only want to take informative examples, we instead choose to select the tth example $x_t = x$ according to a probability $P(x)$ so that our final labeled samples are coming from the distribution λ where $\lambda(x) \propto P(x)\nu(x)$. In particular, $P(x)$ should be chosen according to the following optimization problem

$$P^* = \text{argmin}_{P : \mathcal{X} \to [0,1]} \tau \mathbb{E}_{X \sim \nu}[P(X)] \quad \text{subject to} \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z_* - z\|^2_{\mathbb{E}_{X \sim \nu}[\tau P(X) XX^\top]^{-1}}}{\langle z_* - z, \theta_* \rangle^2} \beta_\delta \leq 1 \tag{4.2}$$

for $\beta_\delta = \log(|\mathcal{Z}|/\delta)$ where the objective captures the number of samples we select using P^*, and the constraint captures the fact that we have solved the problem. Remarkably, we can reparametrize this result in terms of an optimization problem over $\lambda \in \Delta_\mathcal{X}$ instead of $P^* : \mathcal{X} \to [0,1]$ as

$$\tau \mathbb{E}_{X\sim\nu}[P^*(X)] = \min_{\lambda\in\triangle_\mathcal{X}} \rho(\lambda)\beta_\delta \quad \text{subject to} \quad \tau \geq \|\lambda/\nu\|_\infty \rho(\lambda)\beta_\delta$$

where $\|\lambda/\nu\|_\infty = \max_{x\in\mathcal{X}} \lambda(x)/\nu(x)$, as shown in Proposition 4. Note that as $\tau \to \infty$ the constraint becomes inconsequential. Also notice that $\rho(\nu)\beta_\delta$ appears to be a necessary amount of labels to solve the problem even if $P(x) \equiv 1$ (albeit, by arguing about minimizing the upperbound of above).

4.2.2 Main results

In this section we formally justify the sketched argument of the previous section, showing nearly matching upper and lower bounds.

Theorem 10 (Lower bound)**.** *Fix any $\delta \in (0,1)$, $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$, and $\theta_* \in \mathbb{R}^d$. Any selective sampling algorithm that is δ-PAC for $\nu \in \triangle_\mathcal{X}$ and terminates after drawing $\mathcal{U}$ unlabelled examples from ν and requests the labels of just $\mathcal{L}$ of them satisfies*

- $\mathbb{E}[\mathcal{U}] \geq \rho(\nu)\log(1/\delta)$*, and*
- $\mathbb{E}[\mathcal{L}] \geq \min_{\lambda\in\triangle_\mathcal{X}} \rho(\lambda)\log(1/\delta)$ *subject to* $\mathbb{E}[\mathcal{U}] \geq \|\lambda/\nu\|_\infty \rho(\lambda)\log(1/\delta)$.

The first part of the theorem quantifies the number of rounds or unlabelled draws $\mathcal{U}$ that *any* algorithm must observe before it could hope to stop and output z_* correctly. The second part describes a trade-off between $\mathcal{U}$ and $\mathcal{L}$. One extreme is if $\mathbb{E}[\mathcal{U}] \to \infty$, which effectively removes the constraint so that the number of observed labels must scale like $\min_{\lambda\in\triangle_\mathcal{X}} \rho(\lambda)\log(1/\delta)$. Note that this is precisely the number of labels required in the pool-based setting where the agent can choose *any* $x \in \mathcal{X}$ that she desires at each time t (e.g. (Fiez et al., 2019)). In the other extreme, $\mathbb{E}[\mathcal{U}] = \rho(\nu)\log(1/\delta)$ so that the constraint in the label complexity $\mathbb{E}[\mathcal{L}]$ is equivalent to $\rho(\nu) \geq \|\lambda/\nu\|_\infty \rho(\lambda)$. This implies that the minimizing λ must either stay very close to ν, or must obtain a substantially smaller value of $\rho(\lambda)$ relative to $\rho(\nu)$ to account for the inflation factor $\|\lambda/\nu\|_\infty$. In some sense, this latter extreme is the most interesting point on the trade-off curve because its asking the algorithm to stop as quickly as the algorithm that observes all labels, but after requesting a minimal number of labels. Note that this lower bound holds even for algorithms that known ν exactly. The proof of Theorem 10 relies on standard techniques from best arm identification lower bounds (see e.g. (Kaufmann et al., 2016; Fiez et al., 2019)).

Remarkably, every point on the trade-off suggested by the lower bound is nearly achievable.

Theorem 11 (Upper bound)**.** *Fix any $\delta \in (0,1)$, $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$, and $\theta_* \in \mathbb{R}^d$. Let $\Delta = \min_{z\in\mathcal{Z}\setminus\{z_*\}} \langle z_* - z, \theta_* \rangle$ and $\beta_\delta \propto \log(\log(\frac{1}{\Delta})|\mathcal{Z}|/\delta)$ where the precise constant is given in the appendix. For any $\tau \geq \rho(\nu)\beta_\delta$ there exists a δ-PAC selective sampling algorithm that observes $\mathcal{U}$ unlabeled examples and requests just $\mathcal{L}$ labels that satisfies with probability at least $1-\delta$*

- $\mathcal{U} \leq \log_2(\frac{4}{\Delta})\,\tau$*, and*
- $\mathcal{L} \leq 3\log_2(\frac{4}{\Delta}) \min_{\lambda\in\triangle_\mathcal{X}} \rho(\lambda)\,\beta_\delta$ *subject to* $\tau \geq \|\lambda/\nu\|_\infty \rho(\lambda)\,\beta_\delta$.

Aside from the $\log(\frac{1}{\Delta})$ factor and the $\log(|\mathcal{Z}|)$ that appears in the β_δ term, this nearly matches the lower bound. Note that the parameter τ parameterizes the algorithm and makes the trade-off between $\mathcal{U}$ and $\mathcal{L}$ explicit. The next section describes the algorithm that achieves this theorem.

4.2.3 Selective Sampling Algorithm

Algorithm 4.1 contains the pseudo-code of our selective sampling algorithm for best-arm identification. Note that it takes a confidence level $\delta \in (0,1)$ and a parameter τ that controls the unlabeled-labeled budget trade-off as input. The algorithm is effectively an elimination style algorithm and closely mirrors the RAGE algorithm for the pool-based setting of best-arm identification problem (Fiez et al., 2019). The key difference, of course, is that instead of being able to plan over the pool of measurements, this algorithm must plan over the x's that the algorithm may *potentially* see and account for the case that it might not see the x's it wants.

Algorithm 4.1. Selective Sampling for Best-arm Identification

1: **Input** $\mathcal{Z} \subset \mathbb{R}^d$, $\delta \in (0,1)$, τ
2: **while** $|\mathcal{Z}_\ell| > 1$ **do**
3: Let $\widehat{P}_\ell, \widehat{\Sigma}_{\widehat{P}_\ell} \leftarrow$OPTIMIZEDESIGN$(\mathcal{Z}_\ell, 2^{-\ell}, \tau)$ // $\widehat{\Sigma}_{\widehat{P}_\ell}$ approximates $\mathbb{E}_{X\sim\nu}[\widehat{P}_\ell(X)XX^\top]$
4: **for** $t = (\ell-1)\tau + 1, \ldots, \ell\tau$ **do**
5: Nature reveals x_t drawn iid from ν (with support $\mathbb{R}^d$)
6: Sample $Q_t(x_t) \sim$ Bernoulli$(\widehat{P}_\ell(x_t))$. If $Q_t = 1$ then observe y_t // $\mathbb{E}[y_t|x_t] = \langle \theta_*, x_t \rangle$
7: Let $\widehat{\theta}_\ell \leftarrow$RIPS$(\{\widehat{\Sigma}_{\widehat{P}_\ell}^{-1} Q_s(x_s) x_s y_s\}_{s=(\ell-1)\tau+1}^{\ell\tau}, \mathcal{Z} \times \mathcal{Z})$ // $\widehat{\theta}_\ell$ approximates θ_*
8: $\mathcal{Z}_{\ell+1} = \mathcal{Z}_\ell \setminus \{z \in \mathcal{Z}_\ell : \max_{z' \in \mathcal{Z}_\ell} \langle z' - z, \widehat{\theta}_\ell \rangle \geq 2^{-\ell}\}$

In round ℓ, the algorithm maintains an active set $\mathcal{Z}_\ell \subseteq \mathcal{Z}$ with the guarantee that each remaining $z \in \mathcal{Z}_\ell$ satisfies, $\langle z_* - z, \theta_* \rangle \leq 8 \cdot 2^{-\ell}$. In each round, on Line 3 of the algorithm, it calls out to a sub-routine OPTIMIZEDESIGN$(\mathcal{Z}, \epsilon, \tau)$ that is trying to approximate the ideal optimal design of (4.2). In particular, the ideal response to OPTIMIZEDESIGN$(\mathcal{Z}, \epsilon, \tau)$ would return a P_ϵ^* and $\Sigma_{P_\epsilon^*} = \mathbb{E}_{X\sim\nu}[P_\epsilon^*(X)XX^\top]$ where P_ϵ^* is the solution to Equation 4.2 with the one exception that the denominator of the constraint is replaced with $\max\{\epsilon^2, \langle \theta_*, z_* - z \rangle^2\}$. Of course, θ_* is unknown so we cannot solve Equation 4.2 (as well as other outstanding issues that we will address shortly). Consequently, our implementation will aim to *approximate* the optimization problem of Equation 4.2. But assuming our sample complexity is not too far off from this ideal, each round should not request more labels than the number of labels requested by the ideal program with $\epsilon = 0$. Thus, the total number of samples should be bounded by the ideal sample complexity times the number of rounds, which is $O(\log(\Delta^{-1}))$. We will return to implementation issues in the next section.

Assuming we are returned $(\widehat{P}_\ell, \widehat{\Sigma}_{\widehat{P}_\ell})$ that approximate their ideals as just described, the algorithm then proceeds to process the incoming stream of $x_t \sim \nu$. As described above, the decision to request the label of x_t is determined by a coin flip coming up heads with probability $\widehat{P}_\ell(x_t)$–otherwise we do not request the label. Given the collected dataset $\{(x_t, y_t, Q_t, \widehat{P}_\ell(x_t))\}_t$, line 7 then computes an estimate $\widehat{\theta}_\ell$ of θ_* using the RIPS estimator of Figure 2.1 (Camilleri et al., 2021a) which will satisfy

$$|\langle z_* - z, \widehat{\theta}_\ell - \theta_* \rangle| \leq O\left(\|z_* - z\|_{\mathbb{E}_{X\sim\nu}[\tau \widehat{P}_\ell(X) XX^\top]^{-1}} \sqrt{\log(2\ell^2|\mathcal{Z}|^2/\delta)} \right) \leq 2^{-\ell}$$

for all $z \in \mathcal{Z}_\ell$ simultaneously with probability at least $1 - \delta$. Thus, the final line of the algorithm eliminates any $z \in \mathcal{Z}_\ell$ such that there exists another $z' \in \mathcal{Z}_\ell$ (think z_*) that satisfies $\langle \widehat{\theta}_\ell, z' - z \rangle > 2^{-\ell}$. The process continues until $\mathcal{Z}_\ell = \{z_*\}$.

4.2.4 Implementation of OPTIMIZEDESIGN

For the subroutine OPTIMIZEDESIGN passed $(\mathcal{Z}_\ell, \epsilon, \tau)$ the next best thing to computing Equation 4.2 with the denominator of the constraint replaced with $\max\{\epsilon^2, \langle\theta_*, z_* - z\rangle^2\}$, is to compute

$$P_\epsilon = \text{argmin}_{P:\mathcal{X}\to[0,1]}\mathbb{E}_{X\sim\nu}[P(X)] \text{ subject to } \max_{z,z'\in\mathcal{Z}_\ell} \frac{\|z - z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon^2}\beta_\delta \le 1 \tag{4.3}$$

and $\Sigma_{P_\epsilon} = \mathbb{E}_{X\sim\nu}[P_\epsilon(X)XX^\top]$ for an appropriate choice of $\beta_\delta = \Theta(\log(|\mathcal{Z}|/\delta))$. To see this, firstly, any $z \in \mathcal{Z}$ with gap $\langle\theta_*, z_* - z\rangle$ that we could accurately estimate would not be included in $\mathcal{Z}_\ell$, thus we don't need it in the $\max$ of the denominator. Secondly, to get rid of z_* in the numerator (which is unknown, of course), we note that for any norm $\max_{z,z'} \|z - z'\| \le \max_z 2\|z - z_*\| \le \max_{z,z'} 2\|z - z'\|$. Assuming we could solve this directly and compute $\Sigma_{P_\epsilon} = \mathbb{E}_{X\sim\nu}[P_\epsilon(X)XX^\top]$, we can obtain the result of Theorem 2 (proven in the Appendix).

However, even if we knew ν exactly, the optimization problem of Equation 4.3 is quite daunting as it is a potentially infinite dimensional optimization problem over $\mathcal{X}$. Fortunately, after forming the Lagrangian with dual variables for each $z - z' \in \mathcal{Z} \times \mathcal{Z}$, optimizing the dual amounts to a finite dimensional optimization problem over the finite number of dual variables. Moreover, this optimization problem is maximizing a simple expectation with respect to ν and thus we can apply standard stochastic gradient ascent and results from stochastic approximation (Nemirovski et al.). Given the connection to stochastic approximation, instead of sampling a fresh $\widetilde{x} \sim \nu$ each iteration, it suffices to "replay" a sequence of $\widetilde{x}$'s from historical data. Summing up, this construction allows us to compute a satisfactory P_ϵ and avoid both an infinite-dimensional optimization problem and requiring knowledge of ν (as long as historical data is available).

Meanwhile, with historical data, we can also empirically compute $\mathbb{E}_{X\sim\nu}[P_\epsilon(X)XX^\top]$. Historical data could mean offline samples from ν or just samples from previous rounds. In this setting, Theorem 2 still holds albeit with larger constants. Theorem 25 in the appendix characterizes the necessary amount of historical data needed. Unfortunately (in full disclosure) the theoretical guarantees on the amount of historical data needed is absurdly large, though we suspect this arises from a looseness in our analysis. Similar assumptions and approaches to historical or offline data have been used in other works in the streaming setting e.g. (Huang et al., 2015).

4.3 Selective Sampling for Binary Classification

We now review streaming Binary Classification in the agnostic setting (Dasgupta et al., 2008; Hanneke et al., 2014; Huang et al., 2015) and show that our approach can be adapted to this setting. Consider a binary classification problem where $\mathcal{X}$ is the example space and $\mathcal{Y} = \{-1, 1\}$ is the label space. Fix a hypothesis class $\mathcal{H}$ such that each $h \in \mathcal{H}$ is a classifier $h : \mathcal{X} \to \mathcal{Y}$. Assume there exists a fixed regression function $\eta : \mathcal{X} \to [0, 1]$ such that the label of x is Bernoulli with probability $\eta(x) = \mathbb{P}(Y = 1|X = x)$. Being in the agnostic setting, we make no assumption on the relationship between $\mathcal{H}$ and η. Finally, fix any $\nu \in \triangle_\mathcal{X}$ and $\pi \in \triangle_\mathcal{X}$. Given known $\mathcal{X}, \mathcal{H}$ and unknown regression function η, at each time $t = 1, 2, \ldots$:

1. Nature reveals $x_t \sim \nu$
2. Player chooses $Q_t \in \{0, 1\}$. If $Q_t = 1$ then nature reveals $y_t \sim \text{Bernoulli}(\eta(x_t)) \in \{-1, 1\}$
3. Player optionally decides to stop at time t and output some $\widehat{h} \in \mathcal{H}$.

Define the *risk* of any $h \in \mathcal{H}$ as $R_\pi(h) := \mathbb{P}_{X\sim\pi, Y\sim\eta(X)}(Y \neq h(X))$. If the player stops at time $\mathcal{U}$ after observing $\mathcal{L} = \sum_{t=1}^{\mathcal{U}} Q_t$ labels, the objective is to identify $h_* = \arg\min_{h\in\mathcal{H}} R_\pi(h)$ with probability at least $1-\delta$ while minimizing a trade-off of $\mathcal{U}, \mathcal{L}$. Note that h_* is the true risk minimizer with respect to distribution π but we observe samples $x_t \sim \nu$; π is not necessarily equal to ν. While we have posed the problem as identifying the potentially unique h^*, our setting naturally generalizes to identifying an ϵ-good h such that $R_\pi(h) - R_\pi(h_*) \leq \epsilon$.

We will now reduce selective sampling for binary classification problem to selective sampling for best arm identification, and thus immediately obtain a result on the sample complexity. For simplicity, assume that $\mathcal{X}$ and $\mathcal{H}$ are finite. Enumerate $\mathcal{X}$ and for each $h \in \mathcal{H}$ define a vector $z^{(h)} \in [0,1]^{|\mathcal{X}|}$ such that $z_x^{(h)} := \pi(x)\mathbf{1}\{h(x) = 1\}$ for $z^{(h)} = [z_x^{(h)}]_{x\in\mathcal{X}}$. Moreover, define $\theta^* := [\theta_x^*]_{x\in\mathcal{X}}$ where $\theta_x^* := 2\eta(x) - 1$. Then

$$\begin{aligned} R_\pi(h) &= \mathbb{E}_{X\sim\pi, Y\sim\eta(X)}[\mathbf{1}\{Y \neq h(X)\}] = \sum_{x\in\mathcal{X}} \pi(x)(\eta(x)\mathbf{1}\{h(x) \neq 1\} + (1-\eta(x))\mathbf{1}\{h(x) \neq 0\}) \\ &= \sum_{x\in\mathcal{X}} \pi(x)\eta(x) + \sum_{x\in\mathcal{X}} \pi(x)(1-2\eta(x))\mathbf{1}\{h(x) = 1\} = c - \langle z^{(h)}, \theta^*\rangle \end{aligned}$$

where $c = \sum_{x\in\mathcal{X}} \pi(x)\eta(x)$ does not depend on h. Thus, if $\mathcal{Z} := \{z^{(h)}\}_{h\in\mathcal{H}}$ then identifying $h_* = \arg\min_{h\in\mathcal{H}} R_\pi(h)$ is equivalent to identifying $z_* = \arg\max_{z\in\mathcal{Z}} \langle z, \theta^*\rangle$. We can now apply Theorem 11 to obtain a result describing the sample complexity trade-off. First define,

$$\rho_\pi(\lambda, \varepsilon) := \max_{z\in\mathcal{Z}\setminus\{z_*\}} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\lambda}[XX^\top]^{-1}}}{\max\{\langle \theta_*, z_* - z\rangle^2, \varepsilon^2\}} = \max_{h\in\mathcal{H}\setminus\{h_*\}} \frac{\mathbb{E}_{X\sim\pi}\left[\mathbf{1}\{h(X) \neq h'(X)\}\frac{\pi(X)}{\lambda(X)}\right]}{\max\{(R_\pi(h) - R_\pi(h^*))^2, \varepsilon^2\}}$$

An important case of the above setting is when $X \sim \nu$ and $\pi = \nu$, i.e. we are evaluating the performance of a classifier relative to the same distribution our samples are drawn from. This is the setting of (Dasgupta et al., 2008; Huang et al., 2015; Hanneke et al., 2014). The following theorem shows that the sample complexity obtained by our algorithm is at least as good as the results they present.

Theorem 12. *Fix any $\delta \in (0,1)$, domain $\mathcal{X}$ with distribution ν, finite hypothesis class $\mathcal{H}$, regression function $\eta : \mathcal{X} \to [0,1]$. Set $\epsilon \geq 0$ and $\beta_\delta = 2048\log(4\log_2^2(4/\epsilon)|\mathcal{H}|/\delta)$. Then for $\tau \geq \rho_\pi(\nu, \epsilon)\beta_\delta$ there exists a selective sampling algorithm that returns $h \in \mathcal{H}$ satisfying $R_\pi(h) - R_\pi(h^*) \leq \epsilon$ by observing $\mathcal{U}$ unlabeled examples and requesting just $\mathcal{L}$ labels such that*

- $\mathcal{U} \leq \log_2(4/\epsilon)\tau$
- $\mathcal{L} \leq 3\log_2(\frac{4}{\varepsilon}) \min_{\lambda\in\triangle_\mathcal{X}} \rho_\pi(\lambda, \varepsilon)\beta_\delta \quad$ *s.t.* $\quad \tau \geq \|\lambda/\nu\|_\infty \rho_\pi(\lambda, \varepsilon)\beta_\delta$

with probability at least $1-\delta$. Furthermore when $\nu = \pi$ and if $\tau \geq 16\rho(\nu, \epsilon)\beta_\delta$ we have that

$$\mathcal{L} \leq 36\log_2(4/\epsilon)\left(\tfrac{R_\nu(h^*)^2}{\epsilon^2} + 4\right) \sup_{\xi\geq\epsilon} \theta^*(2R_\nu(h^*) + \xi, \nu)\beta_\delta$$

where $\theta^*(u, \nu)$ is the disagreement coefficient, defined in Appendix C.5.

Note that if τ is sufficiently large then the labeled sample complexity we obtain $\min_{\lambda\in\Delta_\mathcal{X}} \rho(\lambda, \epsilon)$ could be significantly smaller than previous results in the streaming setting, e.g. see (Katz-Samuels et al., 2021). The proof of Theorem 12 can be found in Appendix C.5.

4.4 Solving the Optimization Problem

Recall that in Algorithm 4.1, during round ℓ, we need to solve optimization problem (4.3). Solving this optimization problem is not trivial because the number of variables can potentially be infinite if $\mathcal{X}$ is an infinite set. In this section, we will demonstrate how to reduce it to a finite-dimensional problem by considering its dual problem. To simplify the notation, let $\mathcal{Y}_\ell = \{z - z' : z, z' \in \mathcal{Z}_\ell, z \neq z'\}$, and rewrite the problem as follows, where $c_\ell > 0$ is a constant that may depend on round ℓ.

$$\begin{array}{rl} \min_P & \mathbb{E}_{X\sim\nu}\left[P(X)\right] \\ \text{subject to} & y^\top \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right]^{-1} y \leq c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell, \\ & 0 \leq P(x) \leq 1, \quad \forall x \in \mathcal{X}. \end{array} \tag{4.4}$$

Using the Schur complement technique, we show in Lemma 47 (Appendix C.3) the following equivalence: $y^\top \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right]^{-1} y \leq c_\ell^2 \iff \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right] \succeq \frac{1}{c_\ell^2} yy^\top$. This transforms a constraint involving matrix inversion into one with ordering between PSD matrices. Then, we remove the bound constraints $0 \leq P(x) \leq 1$, $\forall x \in \mathcal{X}$ by introducing the barrier function $-\log(1-x) - \log(x)$. That is, instead of working with the objective $\mathbb{E}_{X\sim\nu}\left[P(X)\right]$ directly, we consider the following problem.

$$\begin{array}{rl} \min_P & \mathbb{E}_{X\sim\nu}[P(X) - \mu_b(\log(1-P(X)) + \log(P(X)))] \\ \text{subject to} & \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right] \succeq \frac{1}{c_\ell^2} yy^\top, \quad \forall y \in \mathcal{Y}_\ell. \end{array} \tag{4.5}$$

Here, $\mu_b \in (0,1)$ is some small constant that controls how strong the barrier is. Intuitively, a smaller μ_b will make problem (4.5) closer to the original problem. We now show that unlike the primal, the dual problem is indeed finite-dimensional. For each constraint of $y \in \mathcal{Y}_\ell$, let the matrix $\Lambda_y \succeq \mathbf{0}$ be its dual variable. Further, let $\Lambda = \sum_{y\in\mathcal{Y}_\ell} \Lambda_y$ and $\mathbf{\Lambda} = (\Lambda_y)_{y\in\mathcal{Y}_\ell}$. The corresponding Lagrangian is

$$\mathcal{L}\left(\mathbf{\Lambda}, P\right) = \mathbb{E}_{X\sim\nu}\left[P(X) - \mu_b\left(\log(1-P(X)) + \log(P(X))\right) - P(X)X^\top \Lambda X\right] + \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y y.$$

The dual problem is $\max_{\Lambda_y\succeq\mathbf{0},\forall y\in\mathcal{Y}_\ell} \min_P \mathcal{L}\left(\mathbf{\Lambda}, P\right)$. Notice that minimization over $P : \mathcal{X} \mapsto [0,1]$ can be done via minimizing $P(x)$ point-wise for each $x \in \mathcal{X}$. To do this, we take the gradient with respect to each $P(x)$ and set it to zero to get

$$1 + \frac{\mu_b}{1-P(x)} - \frac{\mu_b}{P(x)} - x^\top \Lambda x = 0. \tag{4.6}$$

Solving this equation and defining $q_\Lambda(x) = x^\top \Lambda x - 1$, we get

$$P_\Lambda(x) = \frac{1}{2} - \frac{\mu_b}{q_\Lambda(x)} + \frac{\sqrt{\left(2\mu_b - q_\Lambda(x)\right)^2 + 4\mu_b q_\Lambda(x)}}{2q_\Lambda(x)}. \tag{4.7}$$

Note that if $\mu_b = 0$ (no barrier), the above reduces to the "threshold" decision rule $P_\Lambda(x) = \frac{1}{2} + \frac{|q_\Lambda(x)|}{2q_\Lambda(x)}$, which gives 0 when $q_\Lambda(x) < 0$ and 1 when $q_\Lambda(x) > 0$.[2] This is exactly the hard elliptical threshold rule mentioned before, in which whether to query the label for x depends on whether it falls inside ($x^\top \Lambda x < 1$) or outside ($x^\top \Lambda x > 1$) of the ellipsoid defined by the positive semidefinite matrix Λ. A visualization of the decision rule P_Λ is given in Figure C.1 in the Appendix.

[2] When $q_\Lambda(x) = 0$, $P_\Lambda(x)$ is undetermined from the dual.

Now, by plugging in $P_\Lambda(x)$, our dual problem becomes $\max_{\Lambda_y \succeq \mathbf{0}, \forall y} D(\mathbf{\Lambda}) := \mathcal{L}(\mathbf{\Lambda}, P_\Lambda)$. This is a finite-dimensional optimization problem, and can be solved by projected gradient ascent (or projected stochastic gradient ascent when we have only samples from ν). The gradient of $D(\mathbf{\Lambda})$ is

$$\begin{aligned}\nabla_{\Lambda_y} D(\mathbf{\Lambda}) &= \mathbb{E}_{X\sim\nu}\left[\left(1+\frac{\mu_b}{1-P_\Lambda(x)}-\frac{\mu_b}{P_\Lambda(X)}-X^\top \Lambda X\right)\nabla_{\Lambda_y} P_\Lambda(X) - P_\Lambda(X) X X^\top\right] + \frac{yy^\top}{c_\ell^2} \\ &= \frac{yy^\top}{c_\ell^2} - \mathbb{E}_{X\sim\nu}\left[P_\Lambda(X) X X^\top\right]. \qquad \text{(Since } P_\Lambda(X) \text{ solves Eq. (4.6))}\end{aligned}$$

The algorithm to solve the problem has been summarized in Algorithm 4.2, in which the gradient during kth iteration is replaced by its unbiased estimator $\frac{yy^\top}{c_\ell^2} - P_{\hat{\Lambda}^{(k)}}(x_k) x_k x_k^\top$. The adaptive learning rate is chosen by following the discussion in chapter 4 of (Orabona, 2019). Optimizing the assignment of $\hat{\Lambda}_y$ to each y in line 9 ensures that the re-scaling step in line 10 increases the function value in an optimized way. Finally, the re-scaling step is used to ensure that the output primal objective value $\mathbb{E}_{X\sim\nu}[P(X)]$ is bounded well, which will be explained in more details in Appendix C.3.

Algorithm 4.2. Projected Stochastic Gradient Ascent to Solve OPTIMIZEDESIGN

1: **Input:** Number of iterations K; number of samples u; barrier weight $\mu_b \in (0,1)$
2: Initialize $\hat{\Lambda}_y^{(0)} = \mathbf{0}$ for each $y \in \mathcal{Y}_\ell$
3: **for** $k = 0, 1, 2, \ldots, K-1$ **do**
4: Sample $x_k \sim \nu$
5: Set $g_{k,y} = \frac{yy^\top}{c_\ell^2} - P_{\hat{\Lambda}^{(k)}}(x_k) x_k x_k^\top$, where P_Λ is defined in Eq. (4.7)
6: Set $\hat{\Lambda}_y^{(k+1)} \leftarrow \hat{\Lambda}_y^{(k)} + \eta_k g_{k,y}$ for each $y \in \mathcal{Y}_\ell$, where $\eta_k = \frac{1}{\sqrt{2\sum_{s=1}^k \sum_{y\in\mathcal{Y}_\ell} \|g_{s,y}\|_2^2}}$
7: Update $\hat{\Lambda}_y^{(k+1)} \leftarrow \Pi_{\mathbb{S}_+^d}(\hat{\Lambda}_y^{(k+1)})$ for each $y \in \mathcal{Y}_\ell$, a projection to the set of $d \times d$ PSD matrices
8: Let $\hat{\Lambda}_y = \frac{1}{K}\sum_{k=1}^K \hat{\Lambda}_y^{(k)}$ for each $y \in \mathcal{Y}_\ell$ and $\hat{\Lambda} = \sum_{y\in\mathcal{Y}_\ell} \hat{\Lambda}_y$
9: Update $(\hat{\Lambda}_y)_{y\in\mathcal{Y}_\ell} \leftarrow \arg\max_{\mathbf{\Lambda}} \sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y y$, subject to $\sum_{y\in\mathcal{Y}_\ell} \Lambda_y = \hat{\Lambda}, \Lambda_y \succeq \mathbf{0}, \forall y \in \mathcal{Y}_\ell$.
10: Find $s^* \leftarrow \arg\max_{s\in[0,1]} D_E(s \cdot \hat{\mathbf{\Lambda}})$, where D_E empirically evaluates D using u i.i.d. samples
11: **return** $\widetilde{\Lambda} = s^* \cdot \sum_{y\in\mathcal{Y}_\ell} \hat{\Lambda}_y$

Let $\mathbf{\Lambda}^*$ be an optimal solution for $D(\mathbf{\Lambda})$. Intuitively, as long as we run this algorithm with sufficiently large number of iterations K and number of samples u, we can guarantee that $D(\widetilde{\mathbf{\Lambda}})$ and $D(\mathbf{\Lambda}^*)$ are close enough with high probability, which in turn guarantees that the primal constraints are violated by only a tiny amount and $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right]$ is close enough to the optimal value. Specifically, we can prove the following theorem.

Theorem 13. *Suppose $\|x\|_2 \le M$ for any $x \in \mathrm{supp}(\nu)$ and $\Sigma = \mathbb{E}_{X\sim\nu}\left[XX^\top\right]$ is invertible. Let $\mathbf{\Lambda}^* \in \arg\max_{\Lambda_y\succeq\mathbf{0},\forall y\in\mathcal{Y}_\ell} D(\mathbf{\Lambda})$ and $\kappa(\Sigma) = \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Sigma)}$ be its condition number. Assume $\|\Lambda^*\|_F > 0$ and define $\omega = \min_{\Gamma\in\mathbb{S}^d:\|\Gamma\|_F=1} \mathbb{E}_{X\sim\nu}\left[\left(X^\top \Gamma X\right)^2\right]$, where $\mathbb{S}^d$ is the set of $d \times d$ symmetric matrices.*

Then, $\Lambda^* = \sum_{y\in\mathcal{Y}_\ell} \Lambda_y^*$ is unique. Further, for any $\epsilon > 0$ and $\delta > 0$, if it holds that $\mu_b \le O\left(\sqrt{\|\Lambda^*\|_F \kappa(\Sigma)} M\right) \cdot \sqrt{(1+\epsilon)/\epsilon}$ and

$$K \ge O\left(\frac{|\mathcal{Y}_\ell|^3 \kappa(\Sigma)^2 \|\Lambda^*\|_F^8 M^{16} \log(1/\delta)}{\omega^2 \mu_b^6}\right) \cdot \left(\frac{1+\epsilon}{\epsilon}\right)^2, u \ge O\left(\frac{\kappa(\Sigma)^2 \|\Lambda^*\|_F^6 M^{16} \log(1/\delta)}{\omega^2 \mu_b^6}\right) \cdot \left(\frac{1+\epsilon}{\epsilon}\right)^2,$$

then, with probability at least $1-\delta$*, Algorithm 4.2 will output* $\widetilde{\Lambda}$ *that satisfies*

- $y^\top \mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)XX^\top\right]^{-1} y \le (1+\epsilon)c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell$.
- $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right] \le \mathbb{E}_{X\sim\nu}\left[\widetilde{P}(X)\right] + 4\sqrt{\mu_b}$*, where* $\widetilde{P}$ *is the optimal solution to problem* (4.4) *with barrier constraint repaced by* $0 \le P(x) \le 1-\mu_b, \forall x \in \mathcal{X}$.

The proof is in Appendix C.3. Although $\widetilde{P}$ is not exactly the same as the optimal solution of the original problem (4.4), when μ_b is sufficiently small, they will be very close. Meanwhile, it should be noted that Theorem 13 mainly reveals that with sufficiently large number of iterations and number of samples, Algorithm 4.2 can output sufficiently good solution. In future work, we plan to examine how much this bound can be improved via a tighter analysis.

Finally, notice that Algorithm 4.2 needs to maintain $|\mathcal{Y}_\ell|\, d^2 = O(|\mathcal{Z}_\ell|^2\, d^2)$ variables, which can be large when we have a large set $\mathcal{Z}_\ell$. Therefore, as an alternative, we also propose Algorithm C.1 that only needs to maintain d^2 variables but requires more computational power in each iteration. The details are given in Appendix C.3.

4.5 Empirical results

In this section we present a benchmark experiment validating the fundamental trade-offs that are theoretically characterized in Theorem 10 and Theorem 11. We take inspiration from (Soare et al., 2014) to define our experimental protocol:

- $d = 2$, a two-dimensional problem.
- $\mathcal{Z} = [\mathbf{e}_1, \mathbf{e}_2, (\cos(\omega), \sin(\omega))]$ for $\omega = 0.3$, where $\mathbf{e}_1, \mathbf{e}_2$ are canonical vectors.
- $\theta_* = 2\mathbf{e}_1$ and $y = x^\top\theta_* + \eta$, where $\eta \sim \mathcal{N}(0,1)$.
- The distribution ν for streaming measurements $x_t \overset{i.i.d.}{\sim} \nu$ is such that $x_t = (\cos(2I_t\pi/N), \sin(2I_t\pi/N))$ where $I_t \in \{0,\dots,N-1\}$, $\mathbb{P}(I_t = i) \propto \cos(2i\pi/N)^2$, and $N = 30$.

In this problem, the angle ω is small enough that the item $(\cos(\omega), \sin(\omega))$ is hard to discriminate from the best item $\mathbf{e}_1$. As argued in (Soare et al., 2014), an efficient sampling strategy for this problem instance would be to pull arms in the direction of $\pm\mathbf{e}_2$ in order to reduce the uncertainty in the direction of interest, $\mathbf{e}_1 - (\cos(\omega), \sin(\omega))$. However, the distribution ν is defined such that it is more likely to receive a vector x_t in the direction of $\pm\mathbf{e}_1$ rather than $\pm\mathbf{e}_2$. Thus, if one seeks a small label complexity, then P should be taken to reject measurements in the direction of $\pm e_1$.

In the benchmark experiment, we compare the following three algorithms which all use Algorithm 4.1 as a meta-algorithm and just swap out the definition of $\widehat{P}_\ell$. `Naive Algorithm` uses no selective sampling so that $\widehat{P}_\ell(x) = 1$ for all x; the `Oracle Algorithm` uses $\widehat{P}_\ell = P_*$ where P_* is the ideal solution to (4.2), and `Our Algorithm` uses the solution to (4.5) for $\widehat{P}_\ell$, where we take $\mu_b = 2\times 10^{-5}$. We swept over the values of τ and plotted on the y-axis the amount of labeled data needed before termination, as shown in Figure 4.1.

We observe in Figure 4.1 that the algorithms using non-naive selection rules require far less label complexity than the naive algorithm for all τ. This reflects the intuition that selection strategies that focus on requesting the more informative streaming measurements are much more efficient than naively observing

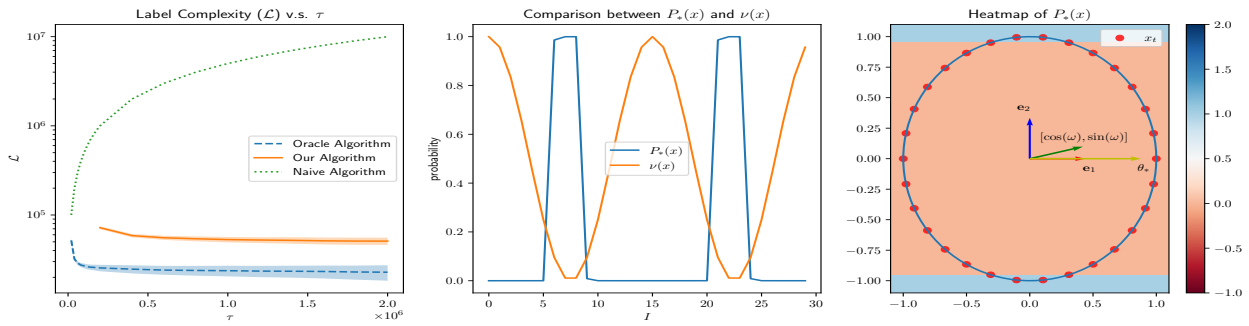


Figure 4.1: (left) For each value of τ, we plot the average label complexity over 50 repeated trials. (middle) Visualization of $P_*(x)$ and $\nu(x)$ v.s. x, where x is indexed by I such that $x_I = (\cos(2I\pi/N), \sin(2I\pi/N))$. Here, P_* is solved with $\tau = 4 \times 10^5$ and distribution ν is not normalized. (right) A heat map of $P_*(x)$ along with the setting of experimental protocol.

every streaming measurement. Meanwhile, the trade-off between label complexity $\mathcal{L}$ and sample complexity $\mathcal{U}$ characterized in Theorem 10 and Theorem 11 is precisely illustrated in Figure 4.1. Indeed, we see the number of labels queried by the two selective sampling algorithms decrease as the number of unlabeled data seen in each round increases.

4.6 Conclusion

In this chapter, we proposed a new approach for the important problem of *selective sampling for best arm identification*. We provide a lower bound that quantifies the trade-off between labeled samples and stopping time and also presented an algorithm that nearly achieves the minimal label complexity given a desired stopping time.

One of the main limitations of the work presented in this chapter is that our approach depends on a well-specified model following stationary stochastic assumptions. In practice, dependencies over time and model mismatch are common. Utilizing the proposed algorithm outside of our assumptions may lead to poor performance and unexpected behavior with adverse consequences. While negative results justify some of the most critical assumptions we make (e.g., allowing the stream x_t to be arbitrary, rather than iid, can lead to trivial algorithms, see Theorem 7 of (Chen et al., 2021)), exploring what theoretical guarantees are possible under relaxed assumptions is an important topic of future work.

Chapter 5

A/B Testing and Best-arm Identification for Linear Bandits with Robustness to Non-Stationarity

5.1 Introduction

Data-driven decision-making and A/B testing enable businesses to evaluate strategies using real-time customer data, offering insights into customer tendencies. As the use of these methods has increased, these technologies are being utilized to determine problems with smaller effect sizes, while also targeting smaller audiences. These two competing trends of smaller effect sizes and smaller sample sizes make it increasingly challenging to obtain statistical significance and correct inference since the absolute number of observations is limited. Consequently, there is a rising trend in using *adaptive* sampling like multi-armed bandits to obtain the same statistical insights using fewer total observations.

However, using adaptive experimentation schemes can come with many pitfalls. Most algorithms that are effective in practice (e.g., Thompson Sampling) are developed with the assumption that the *environment is stationary* and that rewards from treatments are stochastic. However in practice this is far from the case. Non-stationarity can be introduced from a variety of sources such as user populations that change from hour to hour, customer preferences which vary over the course of a year, changes in one part of a platform that lead to latency and higher bounceback, site-wide promotions and sales, interference from competitors, macroeconomic shifts, and many other disruptions. Many of these issues are often totally unobservable, and therefore cannot be controlled, modeled, or accounted for by an experimenter. Under such an environment, it is also possible for the underlying performance of treatments to wildly change, and as a result, the treatment that is best performing on any given day may change. This makes the concept of "the best-performing arm" poorly defined.

Instead, in time-varying settings, the goal of an experimenter is to identify the "counterfactual best treatment" at the end of the experimentation period. That is, the treatment that would have received the *highest total reward had received all the samples*. However, in the absence of being able to predict or model time-variation, predicting precisely how a treatment would behave at every time point, at which time at most one treatment can be evaluated, is impossible. Fortunately, randomization is a powerful tool to provide the next best thing: unbiased *estimates* of how a treatment would behave as if it had been used at every time in the past. These methods are well-understood in the causal-inference and online learning literature and are commonly known as inverse-propensity score (IPS) estimators. The idea is simple: consider a sequence of

evaluations from n treatments at each time $\{x_t\}_{t=1}^T \subset \mathbb{R}^n$. Note that a procedure can only observe at most one treatment per time denoted as $I_t \in [n]$, which is drawn from a distribution p_t over the n treatments. Then $\widehat{X}_i = \frac{1}{T}\sum_{t=1}^T \frac{\mathbf{1}\{I_t=i\}}{p_{t,i}} x_{t,i}$ is an unbiased estimator of the cumulative gain $\frac{1}{T}\sum_{t=1}^T x_{t,i}$ by

$$\mathbb{E}\left[\frac{\mathbf{1}\{I_t=i\}}{p_{t,i}}x_{t,i}\right] = \sum_{j=1}^n \mathbb{P}(I_t=j)\frac{\mathbf{1}\{j=i\}}{p_{t,i}}x_{t,i} = \sum_{j=1}^n p_{t,j}\frac{\mathbf{1}\{j=i\}}{p_{t,i}}x_{t,i} = x_{t,i}, \tag{5.1}$$

as long as $\min_{t,i} p_{t,i} > 0$. Of course, there is no free lunch, and the variance of $\widehat{X}_i$ behaves like $\frac{1}{T^2}\sum_{t=1}^T 1/p_{t,i}$. Intuitively, to maximize efficiency of the samples we do take for inference, we should try to minimize the probabilities on poor performing treatments and prioritize mass for the high performing treatments. However, if the treatment performances vary over time, it can be challenging to determine how one might do this optimally. Fortunately, Abbasi-Yadkori et al. (2018) proposes a novel solution to defining these probabilities in a dynamic way that achieves a "Best of Both Worlds" (BOBW) guarantee, which is an algorithm called P1 that manages to achieve near-optimal rates regardless of whether the environment is stochastic or arbitrarily non-stationary (adversarial). This seminal work is the gold standard for A/B testing in unpredictable non-stationary settings.

If the number of treatments is small (<10 in practice), BOBW provides a robust solution for practitioners. However, there are many situations that practitioners are interested in for which the number of treatments is very large and intractable for traditional A/B testing. For example, multivariate testing Hill et al. (2017) aims to identify not just a single best item, but a set or bundle of items, such as the best 6 pieces of content to highlight on a home screen. Given n possibilities, this results in $\binom{n}{6}$ total distinct treatments for the A / B test! Given this combinatorial explosion, practitioners have made structural parametric assumptions, such as the expected value of a set of items behaves like

$$\theta^{(0)} + \sum_{i=1}^n \theta_i^{(1)}\alpha_i + \sum_{i=2}^n \sum_{j<i} \theta_{i,j}^{(2)}\alpha_i\alpha_j,$$

where $\alpha \in \{0,1\}^n$ with $\sum_i \alpha_i = 6$ indicates whether an item was included in the set or not. Note that these sums can be succinctly written as $\langle x, \theta\rangle$ for $\theta = (\theta^{(0)}, \theta^{(1)}, \theta^{(2)})^\top \in \mathbb{R}^{1+n+\binom{n}{2}}$ and an appropriate $x \in \{0,1\}^{1+n+\binom{n}{2}}$. This can reduce the overall number of unknowns, and dimension, to just $O(n^2)$ versus $O(n^6)$. But now the vectors $x \in \mathcal{X}$, each associated with a particular bundle, are overlapping and can share information. A similar situation arises if we have features or covariates that describe each possible treatment. For example, a particular song comes with lots of metadata including artist, genre, beats per minute, etc. which can encode the useful properties about the song.

In these kinds of scenario—whether it be multivariate testing or items with feature descriptions—we would like to perform adaptive experimentation in the presence of time-variation. Recall that without covariates, we have solutions like P1 that are near-optimal for time-variation. And without time-variation, there are many methods that take covariates into account and are known to be near-optimal. The work presented in this chapter aims to develop an algorithmic framework for handling covariates with time variation.

The remainder of the chapter is organized as follows. We discuss the related work in Section 5.2 and presents detailed problem formulations in Section 5.3. In Section 5.4, we propose a simple algorithm for general non-stationary environments and then in Section 5.5, we propose a robust algorithm that can simultaneously tackle stationary and non-stationary environments. Experiment results are presented in Section 5.6 and our conclusions in Section 5.7.

5.2 Related Work

The problem of identifying the best arm in linear bandits is a well-established and extensively researched problem. (Soare et al., 2014; Karnin, 2016; Xu et al., 2018; Fiez et al., 2019; Katz-Samuels et al., 2020; Degenne et al., 2020; Jedra and Proutiere, 2020; Wagenmaker and Foster, 2023). Notably, Katz-Samuels et al. (2020); Azizi et al. (2021); Yang and Tan (2021) focus on the fixed-budget setting and are closely related to our work. One notable limitation of these algorithms is their reliance on (unrealistic) stationary settings, which leads to their critical failure when applied in non-stationary scenarios. This motivated increasing interest in studying models for non-stationarity in bandits problems and algorithms agnostic to non-stationary settings, which we review next.

Models for non-stationarity in bandits. A reasonable approach in bandit problems with distribution shifts is to provide tight models for unknown variations in the reward distribution. Most literature in this setting focuses on minimizing the dynamic regret, which compares the reward obtained against the reward of the best arm in each round t. (Garivier and Moulines, 2011) demonstrates that existing methods such as Auer et al. (2002b) could achieve a dynamic regret of $\widetilde{O}(\sqrt{LT})$ when L, the number of distribution shifts, is known. Then, Auer et al. (2019) makes a significant advancement by introducing an adaptive approach with the same dynamic regret but without the knowledge of L. More recently, (Chen et al., 2019; Wei and Luo, 2021) establish analogous results in the contextual bandits settings. Measures of non-stationarity other than L are also considered. In particular, Chen et al. (2019) measures the non-stationarity by total variation and Suk and Kpotufe (2021) proposes the novel notion of severe shifts. Note importantly that while this extensive body of work focuses on building tight models of non-stationarity and developing regret minimization algorithms tuned to them, our work is agnostic to such models.

Agnostic non-stationary bandits (Best of both worlds). Bubeck and Slivkins (2012); Seldin and Slivkins (2014); Seldin and Lugosi (2017); Auer and Chiang (2016); Abbasi-Yadkori et al. (2018); Lee et al. (2021) focus on the "best of both worlds" (BOBW) problem: design a bandit algorithm that agnostically achieves optimal performance in both stationary and non-stationary scenarios, even without prior knowledge of the environment. While most BOBW work focus on regret minimization goals, Abbasi-Yadkori et al. (2018) focuses on BOBW for best-arm identification. In this work, as in Abbasi-Yadkori et al. (2018), we focus on the agnostic setting.

A/B testing. As discussed in the introduction, our work is closely related to non-stationary A/B testing. In settings with non-stationarity and adaptive sample allocations, non-stationarity can lead to Simpson's paradox if the sample means are used to estimate arm means Kohavi and Longbotham (2011). It is common in large-scale industrial platforms to assume that means vary smoothly (Wu et al., 2022), or that the differences between them are constant; i.e., all arms are subject to the same random exogeneous shock (Optimizely, 2023). The recent work Qin and Russo (2022) models time-variation as arising from confounding due to a context distribution and aims to find the arm with the best reward on average under this context distribution. Their goal is similar to ours, but, unlike them, we do not assume a context distribution.

5.3 Preliminaries

Notation. Let $[a : b] = \{a, a+1, \ldots, b\}$ for $a, b \in \mathbb{N}$ with $b > a$ and $[a] = \{1, \ldots, a\}$. For a vector $x \in \mathbb{R}^d$ and symmetric positive semi-definite (PSD) matrix $A \in \mathbb{S}_+^d$, we use $\|x\|_A = \sqrt{x^\top A x}$ to denote the Mahalanobis norm. For a finite set $\mathcal{X} \subset \mathbb{R}^d$ and distribution $\lambda \in \triangle_{\mathcal{X}}$ over $\mathcal{X}$, we use $A(\lambda) = \mathbb{E}_{x\sim\lambda}\left[xx^\top\right]$ to denote the covariance matrix under λ.

5.3.1 Linear Bandits Problem Formulation

General stationary/non-stationary environments. In our work, we assume a standard stationary/non-stationary linear bandits model with fixed horizon T. In particular, let $\mathcal{X} \subset \mathbb{R}^d$ be a finite arm set with $|\mathcal{X}| = K$ such that $\text{span}(\mathcal{X}) = \mathbb{R}^d$. At each time $t = 1, \ldots, T$, the learner will pick some arm $x_t \in \mathcal{X}$ and receive some noisy reward $r_t = x_t^\top \theta_t + \epsilon_t$, where $\epsilon_t \in [-1, 1]$ is some independent zero-mean noise. All parameters $\{\theta_t\}_{t=1}^T$ are chosen and fixed by the environment before the game starts.[1] The ultimate goal of the learner is to find the optimal arm $\arg\max_{x \in \mathcal{X}} x^\top \overline{\theta}_T$, where $\overline{\theta}_T = \frac{1}{T}\sum_{t=1}^T \theta_t$ is the average parameter. This protocol is summarized in Figure 5.1.

Input: time horizon, T; arm set, $\mathcal{X} \subset \mathbb{R}^d$
For $t = 1, \ldots, T$
    The learner plays arm $x_t \in \mathcal{X}$
    The learner receives reward $r_t = x_t^\top \theta_t + \epsilon_t$, where ϵ_t is independent zero-mean noise
The learner recommends arm x_{J_T}

Figure 5.1: General protocol of fixed-budget best-arm identification (BAI) for linear bandits.

For simplicity, we further assume that $\forall t \in [T]$, $\forall x \in \mathcal{X}$, $x^\top \theta_t \in [-1, 1]$ and the optimal arm $\arg\max_{x \in \mathcal{X}} x^\top \overline{\theta}_T$ is unique. Meanwhile, similar to Abbasi-Yadkori et al. (2018), we use the subscript (k) to denote the index of k-th best arm in $\mathcal{X}$, which means to have $x_{(1)}^\top \overline{\theta}_T > x_{(2)}^\top \overline{\theta}_T \geq \cdots \geq x_{(K)}^\top \overline{\theta}_T$. For each arm $k \in [K]$, we define its gap Δ_k as

$$\Delta_k = \begin{cases} (x_{(1)} - x_k)^\top \overline{\theta}_T & \text{if } k \neq (1), \\ (x_{(1)} - x_{(2)})^\top \overline{\theta}_T & \text{if } k = (1). \end{cases}$$

That is,we have $\Delta_{(1)} = \Delta_{(2)} \leq \Delta_{(3)} \leq \cdots \leq \Delta_{(K)}$. As a slight abuse of notation, for unindexed arm $x \in \mathcal{X}$, we will use Δ_x to denote the gap of x. The performance of the learner is measured by its error probability $\mathbb{P}_{\overline{\theta}_T}(J_T \neq (1))$, where J_T is the index of the learner's recommendation and the probability measure is taken over the randomness inside the learner and the reward noise. Finally, we note that when the setting is stationary, we simply have $\theta_1 = \cdots = \theta_T = \theta^*$ and everything else is then defined accordingly.

Remark (Comparison to the adversarial setting)**.** *The traditional oblivious adversarial setting can be viewed as a special case of our non-stationary setting, in which we simply pick $\epsilon_t = 0$ for all t (Abbasi-Yadkori et al., 2018).*

5.3.2 BAI for Linear Bandits in Stationary Environments

In this section, we briefly review the well-studied best-arm identification problem for linear bandits in stationary settings. This problem's complexity, first proposed in Soare et al. (2014), is defined as

$$\rho^*(\theta) = H_{\mathsf{LB}}(\theta) = \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \frac{\left\|x - x_{(1)}\right\|^2_{A(\lambda)^{-1}}}{\Delta_x^2}, \tag{5.2}$$

[1] Theoretically, this non-stationary setting has no essential difference with the adversarial setting. We choose this non-stationary setting mainly to keep our presentation concise.

where the optimal arm index (1) and gaps Δ_k are defined based on the input parameter θ. As discussed in Soare et al. (2014), this complexity is approximately equal to the number of samples required (up to logarithmic terms) to find the best arm by running an oracle algorithm. Later in Fiez et al. (2019), this complexity is proved to be the optimal sample complexity that a BAI algorithm can possibly achieve in a fixed-confidence setting. Recently, Katz-Samuels et al. (2020) proposes algorithm Peace in fixed-budget setting that achieves error probability $\mathbb{P}_\theta\left(J_T \neq (1)\right) \leq \widetilde{O}\left(\exp\left(-\frac{T}{\rho^*(\theta)\log(d)}\right)\right)$.[2]

5.4 Best Arm Identification for Linear Bandits in General Non-Stationary Environments

In this section, we present a simple algorithm G-BAI for the general non-stationary environment and analyze its theoretical guarantee. The algorithm is based on the G-optimal design, which refers to the distribution $\lambda^* \in \triangle_{\mathcal{X}}$ such that

$$\lambda^* = \arg\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \in \mathcal{X}} \|x\|^2_{A(\lambda)^{-1}}. \tag{5.3}$$

Intuitively, G-optimal design allows us to estimate unknown parameter θ_t uniformly well over all directions of the arms in $\mathcal{X}$ (Soare et al., 2014). which is suitable for addressing non-stationarity since θ_t may change arbitrarily and each $x \in \mathcal{X}$ may become the optimal at time t. Meanwhile, to make sure the estimation of θ_t is unbiased in a non-stationary environment, we use an IPS estimator.

Therefore, briefly speaking, at each time t, G-BAI simply samples x_t based on G-optimal design and estimate θ_t through an IPS estimator, whose details are summarized in Algorithm 5.1.[3]

Algorithm 5.1. G-optimal Best-arm Identification (G-BAI)

1: **Input** budget, $T \in \mathbb{N}$; arm set $\mathcal{X} \subset \mathbb{R}^d$
2: Compute G-optimal design λ^* based on Eq. (5.3)
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Sample $x_t \sim \lambda^*$ and receive reward r_t
5: Estimate $\widehat{\theta}_T \leftarrow \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{x\sim\lambda^*}\left[xx^\top\right]^{-1} x_t r_t$
6: **Return** $\arg\max_{x\in\mathcal{X}} x^\top \widehat{\theta}_T$

By the famous Kiefer-Wolfowitz theorem, an important property of the G-optimal design is that $\max_{x\in\mathcal{X}} \|x\|^2_{A(\lambda^*)^{-1}} = d$ (Lattimore and Szepesvári, 2020). With this property, the variance of estimator $\widehat{\theta}_t$ can be easily controlled. We can then bound the error probability of G-BAI by this fact and the result is summarized in the following theorem.

Theorem 14 (Error probability of G-BAI)**.** *Fix time horizon T, arm set $\mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = K$ and arbitrary unknown parameters $\{\theta_t\}_{t=1}^T$. If we run Algorithm 5.1 in this non-stationary environment and obtain x_{J_T}, then it holds that*

$$\mathbb{P}_{\overline{\theta}_T}\left(J_T \neq (1)\right) \leq K\exp\left(-\frac{T}{12H_{\text{G-BAI}}\left(\overline{\theta}_T\right)}\right), \quad \textit{where } H_{\text{G-BAI}}\left(\overline{\theta}_T\right) = \frac{d}{\Delta^2_{(1)}}.$$

[2]Rigorously speaking, the error probability of Peace contains another complexity term called $\gamma^*(\theta)$, which is defined as the minimum of a Gaussian width term. However, as argued in Katz-Samuels et al. (2020), $\gamma^*(\theta)$ is roughly in a same order of $\rho^*(\theta)$.

[3]We can see $\widehat{\theta}_T$ exactly becomes the more commonly-seen IPS estimator examined in Eq. (5.1) if we apply it to the multi-armed bandits setting, in which we have $K = d$ arms and $\mathcal{X} = \{\mathbf{1}_1, \ldots, \mathbf{e}_d\}$.

The proof of Theorem 14 is deferred to Appendix D.2. Here, we briefly compare this result with the one in multi-armed bandits, which can be treated as a special case of linear bandits by taking $\mathcal{X} = \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$ to be the canonical vectors (standard basis) with $K = d$.

In particular, Abbasi-Yadkori et al. (2018) shows that in multi-armed bandits setting, a simple uniform sampling algorithm reaches complexity $H_{\text{UNIF}}\left(\overline{\theta}_T\right) = \frac{K}{\Delta_{(1)}^2}$ and it is optimal in non-stationary environments. Meanwhile, based on Theorem 14, we can see the complexity of G-BAI is $H_{\text{G-BAI}}\left(\overline{\theta}_T\right) = \frac{d}{\Delta_{(1)}^2}$, which is exactly $H_{\text{UNIF}}(\overline{\theta}_T)$ if we treat multi-armed bandits as a special case of linear bandits since $d = K$. Furthermore, if we directly apply G-BAI to multi-armed bandits, meaning to use $\mathcal{X} = \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$, then λ^* is exactly the uniform distribution over $\mathcal{X}$. That is, in multi-armed bandits, G-BAI exactly recovers the optimal complexity in non-stationary environments, which shows that G-BAI is minimax optimal for linear bandits.

5.5 A Robust Algorithm for Stationary/Non-Stationary Environments

In this section, we present and analyze a new robust linear bandits BAI algorithm called P1-RAGE, which performs comparable to G-BAI in non-stationary environments but much better than it in stationary environments. We will show that it attains good error probability in both stationary and non-stationary environments simultaneously, without knowing a priori which environment it will encounter. We first discuss some intuitions behind the algorithm design.

Stationary environments. The development of our algorithm P1-RAGE is largely inspired by the high-level idea of the robust algorithm P1, proposed in Abbasi-Yadkori et al. (2018), and the allocation strategy of RAGE, proposed in Fiez et al. (2019). In particular, as discussed in Abbasi-Yadkori et al. (2018), in multi-armed bandits, to minimize the error probability in stationary environment, we need to control the estimation variance of the optimal arm well enough. Therefore, at each time step, algorithm P1 pulls the current estimated best arm with the highest probability (unnormalized "probability one"), then subsequently the second best arm with second highest probability (unnormalized "probability half") and so on. We can notice that it actually matches the allocation strategy of the successive halving algorithm in Karnin et al. (2013), which is proved to be near-optimal for BAI in stationary multi-armed bandits. Therefore, we design our probability allocation based on the allocation strategy of RAGE, which is proven to be near-optimal for fixed-confidence BAI in stationary linear bandits (Fiez et al., 2019). In particular, with the estimated parameter $\widehat{\theta}_t$, we first find the estimated best arms $\hat{x}_t^* = \arg\max_{x \in \mathcal{X}} x^\top \widehat{\theta}_t$. Then, we use a subroutine to repeatedly and virtually eliminate arms with estimated gaps larger than certain threshold and compute $\mathcal{XY}$-allocation of the (virtually) remaining arms.[4] Then, we average over the allocation probabilities computed during each iteration.

Non-stationary environments. Finally, to address the potential non-stationarity in environments, we uniformly mix the allocation probability computed above with a G-optimal design. With such a mixture, the variance over all arms can be controlled well and thus the algorithm will be robust for both stationary and non-stationary environments. The details of P1-RAGE are summarized in Algorithm 5.2 and the subroutine to compute the allocation probability, called RAGE-Elimination, is summarized in Algorithm 5.3.

We bound the error probability of P1-RAGE under both stationary and non-stationary settings in the following theorem and its proof is deferred to Appendix D.3.

[4]The elimination is virtual because no samples are collected during the elimination subroutine.

Algorithm 5.2. P1-RAGE

Input: budget, $T \in \mathbb{N}$; arm set $\mathcal{X} \subset \mathbb{R}^d$; maximum number of virtual phases, m
Compute G-optimal design λ^* based on Eq. (5.3) and initialize $\lambda_1 = \lambda^*$
for $t = 1, 2, \ldots, T$ **do**
    Sample $x_t \sim \lambda_t$ and receive reward r_t
    Estimate $\widehat{\theta}_t \leftarrow \frac{1}{t} \sum_{s=1}^{t} \mathbb{E}_{x \sim \lambda_s} \left[x x^\top \right]^{-1} x_s r_s$
    Update $\lambda_{t+1} \leftarrow$RAGE-Elimination$(\widehat{\theta}_t, m)$
Return $\arg\max_{x \in \mathcal{X}} x^\top \widehat{\theta}_T$

Algorithm 5.3. RAGE-Elimination

Input: arm set $\mathcal{X} \subset \mathbb{R}^d$; current estimate $\widehat{\theta}_t$; maximum number of virtual phases, m
Find $\hat{x}_t^* \leftarrow \arg\max_{x \in \mathcal{X}} x^\top \widehat{\theta}_t$
Initialize $\mathcal{X}_t^{(0)} \leftarrow \mathcal{X}$ and $i \leftarrow 0$
while $|\mathcal{X}_t^{(i)}| > 1$ and $i \leq m$ **do**
    $\lambda_t^{(i)} \leftarrow \arg\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x, x' \in \mathcal{X}_t^{(i)}} \|x - x'\|^2_{A(\lambda)^{-1}}$
    $\mathcal{X}_t^{(i+1)} \leftarrow \left\{ x \in \mathcal{X}_t^{(i)} \;\middle|\; \widehat{\theta}_t^\top (\hat{x}_t^* - x) \leq 2^{-i} \right\}$
    $i \leftarrow i + 1$
Return $(\bar{\lambda}_t + \lambda^*)/2$, where $\bar{\lambda}_t = \frac{1}{i} \sum_{i'=0}^{i-1} \lambda_t^{(i')}$

Theorem 15 (Error Probability of P1-RAGE)**.** *Fix arm set $\mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = K$ and budget T. For a stationary environment with unknown parameter θ, if $m \geq i_0 = \lceil \log_2 \left(1/\Delta_{(1)}\right) \rceil + 1$, then there exists absolute constant $c > 0$ such that the error probability of P1-RAGE satisfies*

$$\mathbb{P}_\theta \left(J_T \neq (1) \right) \leq 2 i_0 K T \exp\left(-\frac{cT}{H_{\text{P1-RAGE}}(\theta)} \right),$$

$$\text{where } H_{\text{P1-RAGE}}(\theta) = \frac{m i_0}{\Delta_{(1)}} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \frac{\left\| x - x_{(1)} \right\|^2_{A(\lambda)^{-1}}}{\Delta_x} + \frac{m\sqrt{d}}{\Delta_{(1)}} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \left\| x - x_{(1)} \right\|_{A(\lambda)^{-1}}. \quad (5.4)$$

For a non-stationary environment with unknown parameter $\{\theta_t\}_{t=1}^T$, there exists absolute constant $c' > 0$ such that the error probability of P1-RAGE satisfies

$$\mathbb{P}_{\bar{\theta}_T} \left(J_T \neq (1) \right) \leq K \exp\left(-\frac{c' T \Delta_{(1)}^2}{d} \right).$$

We can immediately see that in non-stationary environments, the error probability of P1-RAGE matches (up to a constant) with G-BAI, showing that P1-RAGE is minimax optimal for linear bandits under non-stationarity. On the other hand, because of the $\frac{1}{\Delta_{(1)}}$ factor, we can see that in stationary environments, $H_{\text{P1-RAGE}}(\theta) \gtrsim H_{\text{LB}}(\theta)$ (defined in Eq. (5.2)), which implies that P1-RAGE is suboptimal in stationary settings. However, this should be expected since even for multi-armed bandits, as proved in Abbasi-Yadkori et al. (2018), it is impossible for an algorithm to achieve $H_{\text{LB}}(\theta)$ while being robust to non-stationarity, let alone linear bandits.

Nevertheless, when applying Theorem 15 to multi-armed bandits ($\mathcal{X} = \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$), as long as we

choose $m \approx i_0$, we can show that (Corollary 5 in Appendix D.3)

$$H_{\text{P1-RAGE}}(\theta) = \widetilde{O}\left(\frac{1}{\Delta_{(1)}} \max_{k\in[K]} \frac{k}{\Delta_{(k)}}\right) = \widetilde{O}\left(H_{\text{BOB}}(\theta)\right),$$

where $H_{\text{BOB}}(\theta)$ is the best-of-both-worlds complexity proposed in Abbasi-Yadkori et al. (2018). In particular, Abbasi-Yadkori et al. (2018) proves that $H_{\text{BOB}}(\theta)$ is the best complexity that any algorithm can possibly achieve if it is constrained to be robust to non-stationarity. That is, again, our algorithm P1-RAGE retains the near-optimal complexity for stationary multi-armed bandits if it is constrained to be robust in non-stationary environments.

Remark. *Here, we do not elaborate the proof details of Theorem 15 mainly because we do not recognize them as widely applicable techniques. However, we do want to emphasize that this proof is by no means a simple extension of the analysis of the algorithm P1 in Abbasi-Yadkori et al. (2018). In particular, our proof uses a different set of virtual events based on the estimated gaps. Meanwhile, the analysis of subroutine RAGE-Elimination is intricately tailored to the unique characteristics of being a virtual elimination strategy, which is not presented in neither RAGE nor P1 (Abbasi-Yadkori et al., 2018; Fiez et al., 2019).*

Theoretical limitations of P1-RAGE. Despite being near-optimal in multi-armed bandits, $H_{\text{P1-RAGE}}(\theta)$ includes an extra low-order term $\frac{m\sqrt{d}}{\Delta_{(1)}} \inf_{\lambda\in\triangle_{\mathcal{X}}} \max_{x\neq x_{(1)}} \left\|x - x_{(1)}\right\|_{A(\lambda)^{-1}}$. This term appears because the Bernstein's inequality requires a bound of the estimator's magnitude, which can be removed if the concentration bound only scales with the estimator's variance. Although this can often be accomplished by using Catoni's robust mean estimator (Wei et al., 2020), it requires a concrete confidence level to be specified before estimation, which is not feasible in our fixed budget setting. Finding an approach to circumvent this difficulty and remove this extra term, or alternatively, demonstrate that it is necessary, is an open question.

Remark. *The question of whether the extra term is removable naturally relates the instance-dependent lower bound of this problem. However, proving an instance-dependent lower bound for our setting requires constructing both stationary and non-stationary counterexamples. This task is thereby more challenging compared to proving an instance-dependent lower bound for the fixed-budget best-arm identification problem in linear bandits within a purely stationary setting, an open question that persists (see Yang and Tan (2022) for a minimax lower bound). We thus leave establishing such instance-dependent lower bounds for future work.*

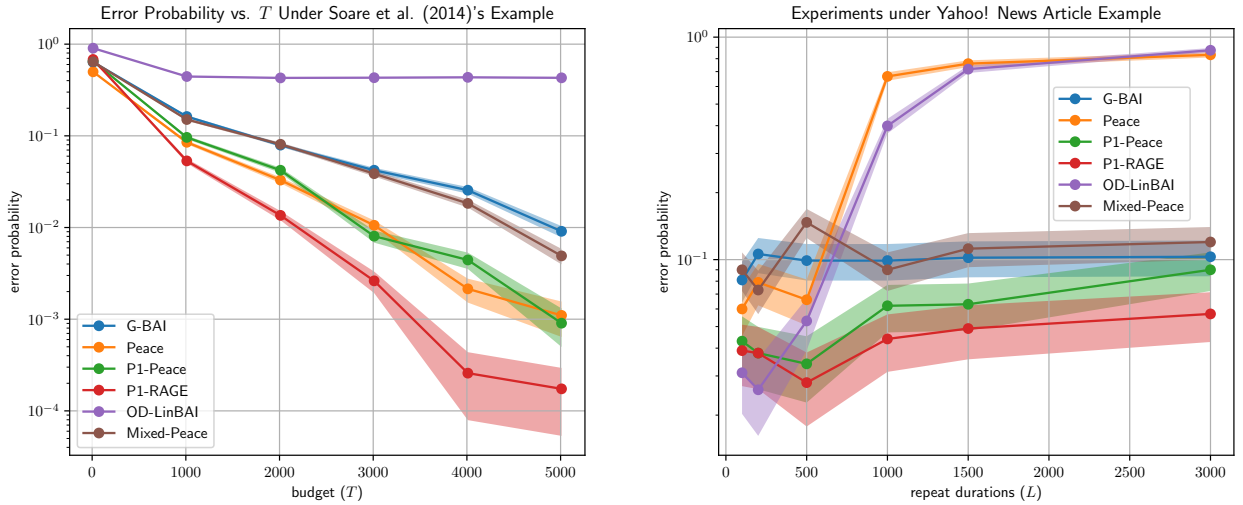
Parameter choice of P1-RAGE. Although P1-RAGE requires a user-specified parameter $m \geq \lceil \log(1/\Delta_{(1)}) \rceil + 1$ to bound the total number of virtual phases, it is not difficult to choose a reasonable value for this parameter in a practical implementation. On the one hand, since its dependence on $\Delta_{(1)}^{-1}$ is only logarithmic, taking some moderate value such as $m = 25$ should safely satisfy $m \geq i_0$ for most practical scenarios; on the other hand, in most real-world applications, a sub-optimal arm should always be acceptable as long as its gap is small enough. Indeed, if we take ϵ to be the largest acceptable sub-optimality gap and take $m \geq \lceil \log(1/\epsilon) \rceil + 1$, then P1-RAGE will output arm x_{J_T} that satisfies $\Delta_{J_T} \leq \max\left\{\epsilon, \Delta_{(1)}\right\}$ with high probability in pure stationary environments (Corollary 6 in Appendix D.3). That is, the output arm will either be an optimal arm if $\epsilon \leq \Delta_{(1)}$ or an arm with an acceptable suboptimality gap ϵ otherwise.

5.6 Experiments

In this section, we present our experiment results on several stationary/non-stationary environments. Since to the best of our knowledge, we are the first to propose best-arm identification algorithms that tackle non-stationarity in linear bandits, the algorithms from other works that we compare with are all specifically

designed for stationary environments. In particular, we will compare our algorithms with Peace, which is the first fixed-budget algorithm for linear bandits and also inspires our algorithmic design (Katz-Samuels et al., 2020), and OD-LinBAI, which is the most recent algorithm of this kind and is claimed to be minimax optimal (Yang and Tan, 2022).

Meanwhile, we also examine two additional heuristically designed algorithms for non-stationary environments. The first one is P1-Peace, which has the same design spirit as P1-RAGE but uses a different Peace-based virtual elimination subroutine; the second one is Mixed-Peace, which is a naive mixture of Peace and the G-optimal design. In particular, while P1-RAGE/P1-Peace combines G-optimal design with what RAGE/Peace would sample *in a full run*, Mixed-Peace simply mixes G-optimal design with what Peace in a stationary environment samples *at each time step*. The details of these two additional algorithms are summarized in Algorithm D.1 and D.3 in Appendix D.1.1, respectively. More implementation details and additional experiments can be found in Appendix D.4.[5]



(a) Each error probability is estimated through at least 2×10^4 independent trials.

(b) Each error probability is estimated through 1000 independent trials.

Figure 5.2: The vertical axis is on log scale and the shaded area represents the 95% confidence interval.

Stationary benchmark example. First, as a sanity check, we consider the famous stationary benchmark example proposed in Soare et al. (2014). In particular, we have $\mathcal{X} = \{\mathbf{e}_1, \ldots, \mathbf{e}_d, x'\}$, where $x' = \cos(\omega)\mathbf{e}_1 + \sin(\omega)\mathbf{e}_2$ with some small $\omega > 0$, and $\overline{\theta}_T = \theta^* = 2\mathbf{e}_1$ so that $x_{(1)} = \mathbf{e}_1$. An efficient algorithm should pick $\mathbf{e}_2$ frequently to reduce the variance in the direction of $\mathbf{e}_1 - x'$. In this example, we pick $d = 10$ and $\omega = 0.1$.

The results are shown in Figure 5.2a. We can see that both our algorithms, P1-RAGE and P1-Peace, perform better than G-BAI and comparably with Peace, showing that our algorithms maintain good performance in stationary environments. Meanwhile, we also notice that Mixed-Peace has performance only comparable to G-BAI, showing that naively mixing the allocation strategy with the G-optimal design can downgrade the performance in stationary environments.

Non-stationary multivariate testing example. We consider a multivariate testing example from Fiez et al. (2019), which is also similar to the one discussed in Introduction. Considering a webpage with D

[5]Code repository is available at `https://github.com/FFTypeZero/bobw_linear`.

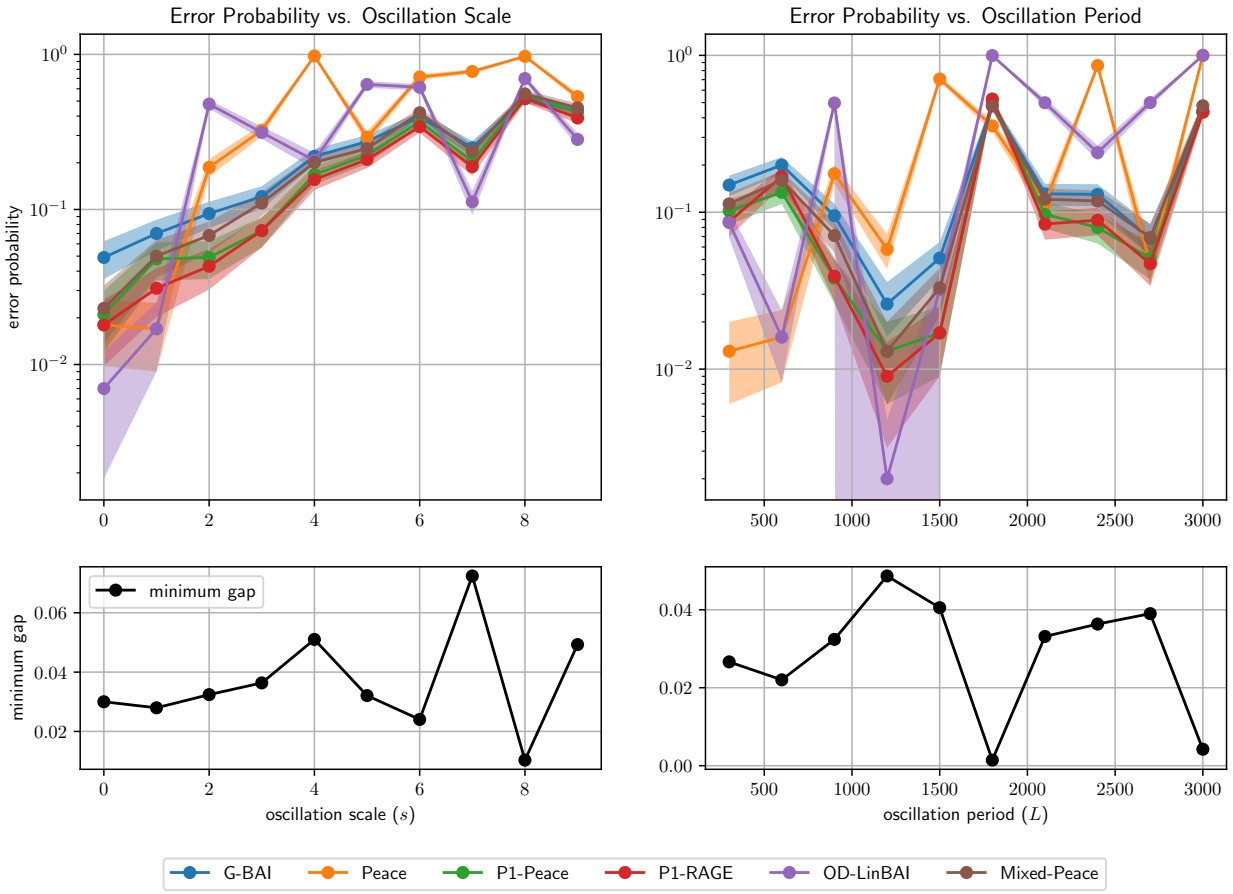


Figure 5.3: Each error probability is estimated through 1000 repeated trials. The bottom two plots give the minimum gap $\Delta_{(1)}$ of each instance as a function of oscillation scale s and oscillation period L.

distinct slots and suppose each slot has two content choices, where we represent each layout as an element $w \in \mathcal{W} = \{-1, 1\}^D$. We hope to maximize the click-through rate and we assume it linearly depends on a layout-determined arm $x \in \mathcal{X}$ in a form of

$$x^\top \theta^* = \theta_0^* + \alpha_1 \sum_{j=1}^{D} \theta_j^* w_j + \alpha_2 \sum_{k=1}^{D-1} \sum_{\ell=k+1}^{D} \theta_{k,\ell}^* w_k w_\ell.$$

Here θ_0^* is the common bias, θ_j^* is the weight of j-th slot and $\theta_{k,\ell}^*$ is the weight of the interaction between k-th and ℓ-th slots. Because of the periodic nature of people's life cycle, it is very likely that the real-world weights will periodically change. Therefore, to construct a non-stationary environment, we randomly oscillate the weights with scale s and period L to get

$$\theta_{t,i} = \theta_i^* + sI \left\|\theta^*\right\|_\infty \sin\left(\frac{2\pi t}{L} + \phi_i\right), \quad \text{where } I \sim \mathsf{Unif}(\{0,1\}), \phi_i \sim \mathsf{Unif}([0, 2\pi]).$$

Here, in the first series of instances, we fix $L = 900$ and take values $s \in \{0, 1, \ldots, 9\}$, and in the second series of instances, we fix $s = 2$ and take values $L \in \{300, 600, \ldots, 3000\}$. Finally, we take $\alpha_1 = 1$, $\alpha_2 = 0.5$, sample each component of θ^* uniformly in $[-0.1, 0.1]$ and guarantee that $\overline{\theta}_T$ has the same optimal arm as θ^*. We take $T = 10^4$ for all settings and the results are shown in Figure 5.3.

From the plots, we can see that the error probabilities of Peace and OD-LinBAI, algorithms designed for stationary environments, can range from near 0 to 1 in different non-stationary environments, which is quite unstable. Meanwhile, we can see that the performance of the other four algorithms, which all in certain way contain a G-optimal design, is relatively much more stable.[6] Furthermore, among these four algorithms, we can see that our algorithms P1-RAGE and P1-Peace consistently outperform (never worse than) G-BAI and Mixed-Peace.

Non-stationary click-through example. To create an instance using real-world data, we use the Yahoo! Webscope Dataset R6A (Yahoo!, 2011).[7] This dataset contains a fraction of user click log of Yahoo!'s news article from May 1st, 2009 to May 10th, 2009. For each click, we take the outer product between user and article features to get a vector in $\mathbb{R}^{36}$ and then we run a principle component analysis to get arm set $\mathcal{Z} \subset \mathbb{R}^{24}$. To create a non-stationary example, we take data from May 1st to May 7th and for each day's data, we fit a ridge regression with regularization 0.01, obtaining $\theta_1^*, \ldots, \theta_7^*$, which can be used to simulate user's weekly periodic behavior. Suppose we receive L visits each day, then, we can define a non-stationary environment where each period consists of $\theta_1^*, \ldots, \theta_1^*, \ldots, \theta_7^*, \ldots, \theta_7^*$ and each θ_i^* repeats for L times. Finally, we form our arm set $\mathcal{X}$ by picking the optimal arm from $\mathcal{Z}$ plus 23 randomly picked arms with gap at least 0.05 so that $\mathrm{span}(\mathcal{X}) = \mathbb{R}^{24}$. We take $T = 2.1 \times 10^4$ and the results are shown in Figure 5.2b. Again, we can see that the performance of Peace and OD-LinBAI is very unstable and the performance of P1-RAGE and P1-Peace consistently outperforms the other two naive G-optimal-design-based algorithms, G-BAI and Mixed-Peace.

5.7 Conclusion and Future Work

We present in this chapter the first two novel robust linear bandits algorithm for fixed-budget best-arm identification, P1-RAGE and P1-Peace, that tackle stationary and non-stationary environments simultaneously while being agnostic to the environment. Theoretically, we prove error probability bounds of P1-RAGE in both stationary and non-stationary environments. Empirically, we show that in stationary settings, both P1-RAGE and P1-Peace perform comparably with algorithms designed for such environments, and in non-stationary settings, they consistently outperform naive algorithms based on G-optimal design.

Finally, several questions still remain open. Is the extra term in $H_{\text{P1-RAGE}}(\theta)$, as discussed in Section 5.5, necessary? What is the optimal complexity for this mixed stationary/non-stationary settings? Answering these questions can serve as promising future directions.

[6] All algorithms fluctuate in the upper right plot mainly because the minimum gaps also have large fluctuation.

[7] `https://webscope.sandbox.yahoo.com/`

Chapter 6

Active Learning with Safety Constraints

6.1 Introduction

In many problems in online decision-making, the goal of the learner is to take measurements in such a way as to learn a near-optimal policy. Oftentimes, though the space of policies may be large, the set of feasible, or safe policies could be much smaller, effectively constraining the search space of the learner. Furthermore, these constraints may themselves depend on unknown problem parameters.

For example, consider the problem of bidding sequentially in a series of auctions where the bidder bids a price w_t, the value of winning an item t is denoted v_t, and the utility of winning that item and paying price p_t is $v_t - p_t$. The goal of the bidder is to choose an optimal strategy amongst bidding strategies $s \in S, s : \mathbb{R} \to \mathbb{R}$. When a bidder is deciding how to choose these strategies, they often face constraints: they may have a budget B they must abide to; they may wish to have those auctions they win be well-distributed across time (e.g. in the case of advertising campaigns); they may want to ensure the set of items they win satisfy some other property (e.g. for advertisements, they might want to ensure they are not over-targeting any demographic group).

As another example, inventory management systems may face similar issues of deciding amongst strategies, where there is some objective function (such as revenue) and a variety of constraints at play in this choice (e.g. capacity of a set of warehouses, employee scheduling constraints, or limits on the duration of delivery lag). They also operate in markets with changing demand and other uncertainties, leading to uncertainty about which strategies are feasible or safe (satisfy constraints) and uncertainty about the revenue they generate.

Both of these scenarios motivate understanding the sample complexity of selecting an action or strategy which approximately maximizes an objective while also satisfying some constraints, where samples are needed to both learn the objective value of actions and whether or not they satisfy said constraints. In this chapter, we study the *active* sample complexity of this task—if the learner can choose which examples to observe and have labeled, how many fewer samples might they need compared to the number needed in a passive setting? We pose this as a best-arm identification problem in the setting of linear bandits with safety constraints, where the goal is to estimate the best arm, subject to it meeting certain (initially unknown) safety constraints. We propose an experiment design-based algorithm which efficiently learns the best safe decision, and show the efficacy of this approach in practice through several experimental examples. To the best of our knowledge, ours is the first approach to handle best-arm identification in linear bandits with safety constraints.

6.1.1 Linear Bandits with Safety Constraints

Let $\delta \in (0,1)$ be a confidence parameter, $\mathcal{X}, \mathcal{Z} \subseteq \mathbb{R}^d$ be finite known sets of vectors, and assume there exists $\theta_* \in \mathbb{R}^d$, $\mu_* \in \mathbb{R}^{m \times d}$ unknown to the learner. For simplicity, we assume that $\|\theta_*\|_2 \leq 1$, and $\|\mu_{*,i}\|_2 \leq 1, i \in [m]$ and $\|x\|_2 \leq 1$, $\|z\|_2 \leq 1, \forall x \in \mathcal{X}, z \in \mathcal{Z}$. The learner plays according to the following protocol: at each time step t the learner chooses some action $x_t \in \mathcal{X}$, observes $(r_t, \{s_{t,i}\}_{i=1}^m)$ where $r_t = \theta_*^\top x_t + w_t^\theta$ and $s_{t,i} = \mu_{*,i}^\top x_t + w_{t,i}^\mu$ for all $i \in [m]$, where $w_t^\theta, w_{t,i}^\mu$ are i.i.d. mean zero 1-subGaussian noise. The choice of action x_t is measurable with respect to the history $\mathcal{F}_t = \{(x_j, r_j, \{s_{j,i}\}_{i=1}^m)\}_{j=1}^{t-1}$. The learner stops at a stopping time τ_δ which is measurable with respect to the filtration generated by $\mathcal{F}_{t \leq \tau}$, and returns $\widehat{z}_\tau \in \mathcal{Z}$. In general, when referring to any expectation $\mathbb{E}$ or probability $\mathbb{P}$, the underlying measure will be with respect to the actions, observed rewards, and internal randomness of the algorithm.

We are interested in the *safe transductive best-arm identification problem* (**STBAI**), where the goal of the learner is to identify

$$z_* := \arg\max_{z \in \mathcal{Z}} z^\top \theta_* \quad \text{s.t.} \quad z^\top \mu_{*,i} \leq \gamma, \forall i \in [m]$$

for some (known) threshold γ. In words, our goal is to identify the best *safe* arm in $\mathcal{Z}$, z_*, where we say an arm z is safe if it satisfies every linear constraint: $z^\top \mu_{*,i} \leq \gamma, \forall i \in [m]$. We are interested in obtaining learners that take the fewest number of samples possible to accomplish this. In practice, we will consider a slightly easier objective. Fix some tolerance $\epsilon > 0$ and let

$$\mathcal{Z}_\epsilon := \{z \in \mathcal{Z} \ : \ z^\top \theta_* \geq z_*^\top \theta_* - \epsilon, \ z^\top \mu_{*,i} \leq \gamma + \epsilon, \forall i \in [m]\}.$$

Then our goal is to obtain an (ϵ, δ)-PAC learner defined as follows:

Definition 3 ((ϵ, δ)-PAC Learner)**.** *A learner is (ϵ, δ)-PAC if for any instance it returns $\widehat{z}_\tau$ such that $\mathbb{P}[\widehat{z}_\tau \in \mathcal{Z}_\epsilon] \geq 1 - \delta$.*

We define the *optimality gap* for any $z \in \mathcal{Z}$ as $\Delta(z) := \theta_*^\top (z_* - z)$, and the *safety gap* for constraint i as $\Delta_{\text{safe}}^i(z) := \gamma - \mu_{*,i}^\top z$. Note that either $\Delta(z)$ or $\Delta_{\text{safe}}^i(z)$ can be negative. If $\Delta(z) < 0$, it follows that z has larger *value*—$z^\top \theta_*$—than the best safe arm z_*, which implies it must be unsafe. If $\Delta_{\text{safe}}^i(z) < 0$ for some i, then arm z is unsafe. We also define the *ϵ-safe optimality gap* as:

$$\Delta^\epsilon(z) = \max_{z' \in \mathcal{Z}} (z' - z)^\top \theta_* \quad \text{s.t.} \quad \min_{i \in [m]} \Delta_{\text{safe}}^i(z') \geq \epsilon. \tag{6.1}$$

$\Delta^\epsilon(z)$ is then the gap in value between arm z and the best arm with minimum safety gap at least ϵ.

Mathematical Notation. Let $\|x\|_A^2 = x^\top A x$ and $\mathfrak{p}(x) := \max\{x, 0\}$. $\widetilde{\mathcal{O}}(\cdot)$ hides factors that are logarithmic in the arguments. $\lesssim$ denotes inequality up to constants. We denote the simplex as $\triangle_{\mathcal{X}} := \{\lambda \in \mathbb{R}_{\geq 0}^{|\mathcal{X}|} : \sum_{x \in \mathcal{X}} \lambda_x = 1\}$.

6.2 Safe Best-Arm Identification in Linear Bandits

6.2.1 Algorithm Definition

The main challenge in algorithm design for the safe best-arm identification problem is ensuring that we are efficiently balancing our exploration between refining our estimates of both the safety gaps, as well as the optimality gaps. Our approach is given in Algorithm 6.1, BESIDE.

Algorithm 6.1. Best Safe Arm Identification (BESIDE)

input: tolerance ϵ, confidence δ
$\iota_\epsilon \leftarrow \lceil \log(\frac{20}{\epsilon}) \rceil$, $\widehat{\Delta}^{i,0}_{\text{safe}}(z) \leftarrow 0, \widehat{\Delta}^{0}(z) \leftarrow 0$ for all $z \in \mathcal{Z}$
for $\ell = 1, 2, \ldots, \iota_\epsilon$ **do**
    $\epsilon_\ell \leftarrow 20 \cdot 2^{-\ell}$
// Phase 1: Solve design to reduce uncertainty in safety constraints
    Define

$$c_\ell(z) = \min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z))$$

    Let τ_ℓ be the minimal value of $\tau \in \mathbb{R}_+$ which is greater than $4 \log \frac{4m|\mathcal{Z}|\ell^2}{\delta}$ such that the objective to the following is no greater than $\epsilon_\ell/100$, and λ_ℓ the corresponding optimal distribution

$$\inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} -\frac{1}{100}\left(c_\ell(z) + \epsilon_\ell\right) + \sqrt{\tau^{-1} \cdot \|z\|^2_{A(\lambda)^{-1}} \cdot \log(\tfrac{4m|\mathcal{Z}|\ell^2}{\delta})}$$

    Sample $x_t \sim \lambda_\ell$, collect τ_ℓ observations $\{(x_t, r_t, s_{t,1}, \ldots, s_{t,m})\}_{t=1}^{\tau_\ell}$
// Phase 2: Estimate safety constraints
    $\{\widehat{\mu}^{i,\ell}\}_{i=1}^m \leftarrow \mathsf{RIPS}(\{(x_t, s_{t,i})\}_{t=1}^{\tau_\ell}, \mathcal{Z}, \frac{\delta}{2m\ell^2})$
    $\widehat{\Delta}^{i,\ell}_{\text{safe}}(z) \leftarrow \gamma - z^\top \widehat{\mu}^{i,\ell} + \|z\|_{A(\lambda_\ell)^{-1}} \sqrt{\tau_\ell^{-1} \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}$
// Phase 3: Refine estimates of optimality gaps
    $\{\widehat{\Delta}^{\ell}(z)\}_{z\in\mathcal{Z}} \leftarrow \textsc{Rage}^\epsilon\Big(\mathcal{Z}, \mathcal{Y}_\ell, \epsilon_\ell, \frac{\delta}{4\ell^2}, \{\widehat{\Delta}_{\text{safe}}(z) \leftarrow \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell}_{\text{safe}}(z))\}_{z\in\mathcal{Z}}\Big)$
// Perform final round of exploration to ensure we find ϵ-good arm
    $\mathcal{Y}_{\text{end}} \leftarrow \{z \in \mathcal{Z} : c_\ell(z) \lesssim \widehat{\Delta}^{i,\ell}_{\text{safe}}(z) + \epsilon\}$
    $\{\widehat{\Delta}^{\text{end}}(z)\}_{z\in\mathcal{Y}_{\text{end}}} \leftarrow \textsc{Rage}^\epsilon(\mathcal{Y}_{\text{end}}, \mathcal{Y}_{\text{end}}, \epsilon, \delta, \{\widehat{\Delta}_{\text{safe}}(z) \leftarrow \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell}_{\text{safe}}(z))\}_{z\in\mathcal{Z}})$
return $\widehat{z} = \text{argmin}_{z\in\mathcal{Y}_{\text{end}}} \widehat{\Delta}^{\text{end}}(z)$

BESIDE relies on a round-based adaptive experimental design approach. In each round BESIDE consists of three phases. In the first phase, it solves an experimental design over $\lambda_\ell \in \triangle_\mathcal{X}$, with the goal of refining our estimates of the safety gaps. It then takes τ_ℓ samples from λ_ℓ. In the second phase these samples are used to estimate the safety constraints, $\widehat{\mu}^{i,\ell}$, and the safety gaps of each arm, $\widehat{\Delta}^{i,\ell}_{\text{safe}}(z)$. Finally, in Phase 3, an additional experimental design is solved which now aims to refine our estimates of the optimality gaps, and the estimates of the optimality gaps $\widehat{\Delta}^{\ell}(z)$ for each $z \in \mathcal{Z}$ are then computed. We encapsulate Phase 3 in a subroutine, $\textsc{Rage}^\epsilon$, which we outline in the following. We now carefully describe each phase—we begin with Phase 2 to explain how our estimator works.

Phase 2: In Phase 2 the algorithm would like to use the τ_ℓ samples drawn from the design λ_ℓ to estimate the constraints for each $z \in \mathcal{Z}$: $z^\top \mu_{*,i}$ for each $i \in [m]$. Past works using adaptive experimental design in the linear bandits literature have utilized the least-squares estimator along with complicated rounding schemes (Fiez et al., 2019) which may require an additional $\text{poly}(d)$ samples each round (this $\text{poly}(d)$ factor could be prohibitively large—for example, in active classification problems, d is the total number of data points). We instead utilize the RIPS estimator described in Chapter 2 (Camilleri et al., 2021a) which gives us a guarantee

of the form: with probability greater than $1-\delta$, for all $z \in \mathcal{Z}$,

$$|z^\top(\widehat{\mu}^{i,\ell} - \mu_{*,i})| \lesssim \|z\|_{A(\lambda_\ell)^{-1}} \cdot \sqrt{\tau_\ell^{-1} \log(\tfrac{4m|\mathcal{Z}|\ell^2}{\delta})}. \tag{6.2}$$

We describe the RIPS estimator in more detail in Section E.2.

Phase 1: By our definition of the experimental design on Line 6, our safety gap estimation error bound in (6.2) satisfies, for each $z \in \mathcal{Z}$:

$$|z^\top(\widehat{\mu}^{i,\ell} - \mu_{*,i})| \lesssim \|z\|_{A(\lambda_\ell)^{-1}} \cdot \sqrt{\tau_\ell^{-1} \log(\tfrac{4m|\mathcal{Z}|\ell^2}{\delta})} \lesssim c_\ell(z) + \epsilon_\ell. \tag{6.3}$$

Note that our design chooses an allocation that minimizes the variance in our estimate of each safety constraint (up to some tolerance), which scales as $\|z\|^2_{A(\lambda)^{-1}}$. This can be thought of as a form of $\mathcal{XY}$*-design*—a design of the form $\inf_{\lambda \in \triangle_\mathcal{X}} \max_{y \in \mathcal{Y}} \|y\|^2_{A(\lambda)^{-1}}$—where here $\mathcal{Y} \leftarrow \mathcal{Z}$ is chosen to reduce our uncertainty in estimating the safety value for each $z \in \mathcal{Z}$. We refer to such a design objective henceforth as $\mathcal{XY}_{\text{safe}}$. Assume that at round $\ell - 1$, we can guarantee

$$\begin{aligned} c_\ell(z) &= \min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell \\ &\lesssim \min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + \mathfrak{p}(\Delta^{\epsilon_{\ell-1}}(z)) + \epsilon_\ell. \end{aligned} \tag{6.4}$$

Then combining the above inequalities, we see that the experiment design on Line 6 aims to minimize the uncertainty in our estimate of $z^\top \mu_{*,i}$ up to a tolerance that scales as the maximum of the four terms in (6.4). It follows that if any of these terms is large, we will only allocate a small number of samples to refining our estimate of arm z. Each one of these terms can be intuitively motivated by thinking through what is needed to prove that an arm $z \neq z_*$.

- z **has small safety gap** $\min_j |\Delta^j_{\text{safe}}(z)|$**:** if this term is large, it implies that minimum safety gap for z is large. To show an arm is safe or unsafe, it suffices to learn each safety gap up to a tolerance a constant factor from its value—regularizing by this term ensures we do just that.

- z **fails some safety constraint** $\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z))$**:** if this term is large, it implies that arm z is very unsafe for some constraint. In this case, we can easily determine z is unsafe, and therefore do not need to reduce our uncertainty in the safety gap any more.

- z **is sub-optimal** $\mathfrak{p}(\Delta^{\epsilon_{\ell-1}}(z))$**:** if this term is large, it implies that z is very suboptimal compared to some safe arm with safety gap at least $\epsilon_{\ell-1}$. In this case, we do not need to estimate z's safety gap, as we will have already eliminated it.

It remains to ensure that (6.4) holds. As we show in Section E.4 through a careful inductive argument, combining (6.3) with our guarantee on the estimates of the optimality gaps obtained in Phase 3, $\widehat{\Delta}^\ell(z)$, is sufficient to guarantee (6.4) holds. In particular, if any gap is greater than ϵ_ℓ it is estimated up to a constant factor, and otherwise it is estimated up to $\mathcal{O}(\epsilon_\ell)$. This ensures that our gaps are estimated at the correct rate while guaranteeing we do not collect too many samples in each round.

Phase 3: In this phase we estimate the suboptimality gaps using RAGE$^\epsilon$. RAGE$^\epsilon$ is inspired by the RAGE algorithm of (Fiez et al., 2019) for best-arm identification. In the interest of space, we defer the full definition of RAGE$^\epsilon$ to Section E.3 but provide some intuition here. After Phase 2, by (6.3) the set of arms $\mathcal{Y}_\ell := \{z \in \mathcal{Z} : c_s(z) \lesssim \widehat{\Delta}^{i,s}(z), \forall i \in [m]\}$ for $s \leq \ell$ are precisely the ones that we can certify are safe (note that we do not need to ever explicitly construct such a set—we can instead maintain an implicit definition through the constraints). RAGE$^\epsilon$ uses an adaptive experimental design procedure to sample in such a way as to optimally estimate the gaps $(z - \widehat{y})^\top \theta_*, \forall z \in \mathcal{Z}$ and some $\widehat{y} \in \mathcal{Y}_\ell$ up to some (sufficient) tolerance. In particular, it also solves an $\mathcal{XY}$-design, but now on the set $\mathcal{Y} \leftarrow \{z - \widehat{y} : z \in \mathcal{Z}\}$. Thus, rather than minimizing $\|z\|^2_{A(\lambda)^{-1}}$, we minimize $\|z - \widehat{y}\|^2_{A(\lambda)^{-1}}$. This design reduces uncertainty on the *differences* between arms, which allows us to refine our estimates of their optimality gaps. Henceforth we refer to such a design as $\mathcal{XY}_{\text{diff}}$. We describe the importance of the choice of design in more detail in Section 6.2.4. Ultimately, if an arm z has value within a factor of ϵ_ℓ of the best safe arm in $\mathcal{Y}_\ell$, and if we have not yet shown arm z is unsafe, then we will estimate its optimality gap up to a constant factor of ϵ_ℓ. If we were maintaining arm sets explicitly (similar to the original RAGE algorithm of (Fiez et al., 2019)) we would eliminate arms at this point.

Remark (Computational Complexity)**.** *The main computational challenge in* BESIDE *and* RAGE$^\epsilon$ *is the calculation of the experimental designs (i.e. Line 6 and the corresponding design in* RAGE$^\epsilon$*). In general, the presence of the square root implies that the resulting optimization problem may not be convex in λ. To handle this issue we note that $2\sqrt{xy} = \min_{\alpha>0} \alpha x + \frac{y}{\alpha}$—thus we can replace the existing design with $\inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} \min_{\alpha > 0} -\frac{1}{100}(c_\ell(z) + \epsilon_\ell) + \alpha \|z\|^2_{A(\lambda)^{-1}} + \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})/(\alpha\tau)$. By appropriately discretizing the space we search over for τ and α we can then apply the Frank-Wolfe algorithm to minimize over λ. While computationally efficient in theory, this procedure is quite complicated and impractical for large problems. In the experiments section we provide a practical heuristic that is motivated by the above algorithm and is computationally efficient for larger problems.*

6.2.2 Main Result

BESIDE achieves the following complexity.

Theorem 16. BESIDE *is (ϵ, δ)-PAC. In other words, with probability at least $1 - \delta$,* BESIDE *returns an arm $\widehat{z} \in \mathcal{Z}$ such that*

$$\widehat{z}^\top \theta_* \geq z_*^\top \theta_* - \epsilon, \quad \min_{i \in [m]} \Delta^i_{\text{safe}}(\widehat{z}) \geq -\epsilon$$

and terminates after collecting at most

$$C \cdot \sup_{\widetilde{\epsilon} \geq \epsilon} \inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} \frac{\|z\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{m|\mathcal{Z}|}{\delta})}{\left(\min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + \mathfrak{p}(\Delta^{\widetilde{\epsilon}}(z)) + \widetilde{\epsilon}\right)^2} \qquad \text{(safety)}$$

$$+ C \cdot \sup_{\widetilde{\epsilon} \geq \epsilon} \inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} \frac{\|z - z_*\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{|\mathcal{Z}|}{\delta})}{\left(\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + \mathfrak{p}(\Delta^{\widetilde{\epsilon}}(z)) + \widetilde{\epsilon}\right)^2} + C_0 \qquad \text{(optimality)}$$

samples for some $C = \operatorname{poly}\log(\frac{1}{\epsilon})$ and $C_0 = \operatorname{poly}\log(\frac{1}{\epsilon}, |\mathcal{Z}|) \cdot \log\frac{1}{\delta}$.

The complexity bound given in Theorem 16 may, at first glance, appear rather opaque, yet it in fact yields a very intuitive interpretation. The first term in the complexity, the safety term, is the complexity

needed to show each arm is safe or unsafe, *if they have not otherwise been eliminated.* As described in the previous section, if $\mathfrak{p}(\Delta^{\widetilde{\epsilon}}(z))$ is large, this implies we have found an arm better than z, so learning its safety value is irrelevant.

The second term in the complexity, the optimality term, corresponds to the difficulty of showing an arm is worse *than the best arm we can guarantee is safe*. Note that we can only guarantee an arm is suboptimal if we can find a safe arm with higher value. Recall the definition of $\Delta^{\widetilde{\epsilon}}(z)$ given in (6.1). Intuitively, $\Delta^{\widetilde{\epsilon}}(z)$ denotes the gap in value between arm z and the best arm with safety gap at least $\widetilde{\epsilon}$. As we make $\widetilde{\epsilon}$ smaller, we can show additional arms are safe, which increases $\Delta^{\widetilde{\epsilon}}(z)$. While this makes it easier to show z is suboptimal, it comes at a cost—the extra samples necessary to decrease our safety tolerance, given by the first term in the complexity. BESIDE trades off between optimizing for each of these terms—gradually decreasing its tolerance on both the safety and optimality terms to more easily eliminate suboptimal arms, while not allocating too many samples to guarantee safety.

To help illustrate this complexity, we consider a simple example with orthogonal arms, i.e. a multi-armed bandit example.

Example 1 (BESIDE on Multi-Armed Bandits)**.** *In the multi-armed bandit setting, we have* $\mathcal{X} = \mathcal{Z} = \{e_1, \dots, e_d\}$*. Let* $m = 1$*,* $d = 3$*, and consider the settings of* θ_* *and* μ_* *given in Figure 6.1. Here we see that arm* e_1 *is safe and has value much higher than any other arm, so* $z_* = e_1$*, and can be shown to be safe relatively easily; arm* e_2 *has near-optimal value but is very unsafe; and arm* e_3 *is unsafe with very small safety gap, but has the smallest value.*

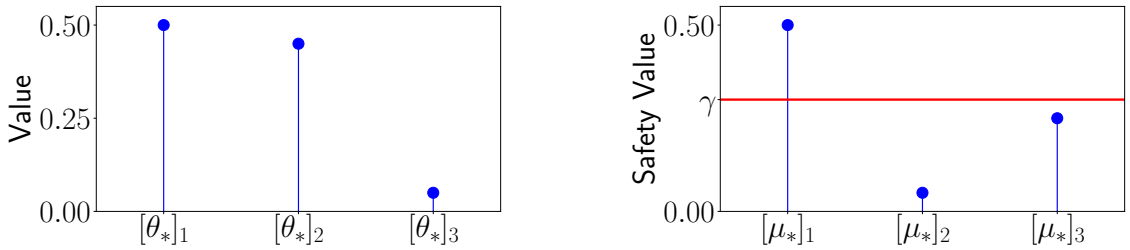


Figure 6.1: Multi-Armed Bandit Instance

Showing e_2 ***is Suboptimal.*** *As* e_2 *has near-optimal value,* $\Delta(e_2)$ *is very small and it is very difficult to show* e_2 *is suboptimal. However,* $-\Delta_{\text{safe}}(e_2) = \mathcal{O}(1)$*, so it is very easy to show* e_2 *is unsafe. It follows that* $\mathfrak{p}(-\Delta_{\text{safe}}(e_2)) = \mathcal{O}(1)$ *so both denominators in our complexity will always be* $\mathcal{O}(1)$ *for* $z = e_2$—BESIDE *does not attempt to show* e_2 *is suboptimal, but instead shows it is unsafe, and therefore does not pay for the small optimality gap of* $\Delta(e_2)$ *in the complexity.*

Showing e_3 ***is Suboptimal.*** *Recall the definition of* $\Delta^{\epsilon}(z) = \max_{z':\Delta_{\text{safe}}(z') \geq \epsilon} \theta_*^\top (z' - z)$*. In this case, for* $\epsilon = \mathcal{O}(1)$*, we will have* $\Delta_{\text{safe}}(e_1) \geq \epsilon$*, which implies that* $\Delta^{\epsilon}(e_3) = \theta_*^\top (e_1 - e_2) = \Delta(e_3) = \mathcal{O}(1)$*. To show* e_3 *is suboptimal, we could either show it is unsafe (which is very difficult) or suboptimal (which is very easy). Observing the sample complexity of Theorem 16, we see that the denominator of both terms will always be* $\mathcal{O}(1)$ *for* $z = e_3$ *since* $\Delta^{\epsilon}(e_2) = \mathcal{O}(1)$—BESIDE *never pays for the small safety gap of* e_3*, it instead takes advantage of the fact that* e_3 *can easily be shown to be suboptimal, and uses this to eliminate it.*

In both of these cases we see that BESIDE *does the "right" thing, always using the easier of the two criteria—either showing an arm is unsafe or suboptimal—to show that* $z \neq z_*$*. Combining the above observations, for* $\epsilon \approx \min\{\Delta(e_3), -\Delta_{\text{safe}}(e_2), \Delta_{\text{safe}}(e_1)\}$*, it follows that on this example the total sample*

complexity of BESIDE *given by Theorem 16 scales as:*

$$\widetilde{\mathcal{O}}\left(\left(\frac{1}{\Delta_{\text{safe}}(e_1)^2}+\frac{1}{\Delta_{\text{safe}}(e_2)^2}+\frac{1}{\Delta(e_3)^2}\right)\cdot\log\frac{1}{\delta}\right)$$

where the $1/\Delta_{\text{safe}}(e_1)^2$ *arises because we must also show* e_1 *is safe.*

6.2.3 Optimality of BESIDE

Optimality in Best-Arm Identification. Consider applying BESIDE to a problem instance where $m = 1$, $\mu_{*,1} = 0$, and $\gamma = 1$. In this case, every arm is safe, and the safety constraints are essentially vacuous—every arm can easily be shown safe. We can therefore think of this as simply an instance of the best-arm identification problem. In this setting, we obtain the following corollary.

Corollary 2. *Consider running* BESIDE *on a problem instance where* $m = 1$*,* $\mu_{*,1} = 0$*, and* $\gamma = 1$*, and set* $\epsilon = \frac{1}{2}\max_{z\neq z_*}\theta_*^\top(z_* - z)$*. Then with probability at least* $1-\delta$*,* BESIDE *returns* z_* *and has sample complexity bounded by:*

$$\widetilde{\mathcal{O}}\left(\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z-z_*\|^2_{A(\lambda)^{-1}}}{\Delta(z)^2}\cdot\log\frac{|\mathcal{Z}|}{\delta}+\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\|z\|^2_{A(\lambda)^{-1}}\cdot\log\frac{|\mathcal{Z}|}{\delta}\right).$$

Up to lower-order terms, this exactly matches the lower bound on best-arm identification given in (Fiez et al., 2019). Thus, in settings where the safety constraint is vacuous, BESIDE hits the optimal rate.

Worst-Case Performance of BESIDE. We next consider the worst-case performance of BESIDE in settings when $\mathcal{X} = \mathcal{Z}$. We have the following result.

Corollary 3. *Assume that* $\mathcal{X} = \mathcal{Z}$*. Then for any* θ_* *and* $(\mu_{*,i})_{i=1}^m$*, the sample complexity of* BESIDE *necessary to return an* ϵ*-good and* ϵ*-safe arm is bounded as* $\widetilde{\mathcal{O}}(\frac{d}{\epsilon^2}\cdot(\log(m|\mathcal{X}|)+\log\frac{1}{\delta}))$.

Theorem 2 of (Wagenmaker et al., 2022a) shows a worst-case lower bound of $\Omega(d^2/\epsilon^2)$ on the sample complexity of identifying an ϵ-optimal arm in the standard linear bandit setting. Safe best-arm identification problems in which the safety constraint is vacuous are at least as hard as the standard best-arm identification problem, since at minimum we need to find the best arm out of every safe arm. Thus, $\Omega(d^2/\epsilon^2)$ is also a worst-case lower bound for the safe best-arm identification problem. The hard instance of (Wagenmaker et al., 2022a) has $|\mathcal{X}| = \mathcal{O}(2^d)$, so it follows that on this instance, BESIDE achieves a complexity of $\widetilde{\mathcal{O}}(\frac{d}{\epsilon^2}\cdot(d+\log\frac{1}{\delta}))$, and therefore BESIDE has optimal dimensionality dependence. In addition, this also implies that safe best-arm identification, in the worst-case, is no harder than the standard best-arm identification problem—it is no harder to find the best *safe* arm, regardless of the number of safety constraints, than to find the best arm, ignoring safety constraints.

6.2.4 The Role of Experiment Design

We can think of the safe best-arm identification problem, in some sense, as an interpolation of the standard best-arm identification problem, as well as the level-set estimation problem, where the goal is to identify $z \in \mathcal{Z}$ satisfying $z^\top\mu_* \leq \gamma$ (Mason et al., 2021). In the former problem, (Fiez et al., 2019) shows that the instance-optimal rate can be attained by running a round-based algorithm and at every round solving an instance of the $\mathcal{XY}_{\text{diff}}$ experiment design, as defined in Section 6.2.1. In the latter problem, (Mason

et al., 2021) also show that a round-based algorithm can hit the instance-optimal rate, but instead solving the $\mathcal{XY}_{\text{safe}}$ problem at each round. It is natural to ask whether either of these strategies could be applied to the safe best-arm identification problem directly, or if it is necessary to alternate between them. The following results show that, on their own, each of these designs is unable to hit the optimal rate.

Proposition 1. *Fix some small enough $\epsilon > 0$. Then there exist instances of the safe best-arm identification problem, $\mathcal{I}_i = (\theta^i_*, \mu^i_*, \mathcal{X}^i, \mathcal{Z}^i)$, $i = 1, 2$, with $d = |\mathcal{X}^i| = |\mathcal{Z}^i| = 2$, $m = 1$, such that:*

- *On $\mathcal{I}^1$, any (ϵ, δ)-PAC algorithm which plays only allocations minimizing $\mathcal{XY}_{\text{diff}}$ must have $\mathbb{E}[\tau_\delta] \geq \Omega\left(\frac{1}{\epsilon^3} \cdot \log \frac{1}{\delta}\right)$, while* BESIDE *identifies an ϵ-optimal arm after $\widetilde{\mathcal{O}}(\frac{1}{\epsilon^2} \cdot \log 1/\delta)$ samples.*
- *On $\mathcal{I}^2$, any (ϵ, δ)-PAC algorithm which plays only allocations minimizing $\mathcal{XY}_{\text{safe}}$ must have $\mathbb{E}[\tau_\delta] \geq \Omega\left(\frac{1}{\epsilon^{3/2}} \cdot \log \frac{1}{\delta}\right)$, while* BESIDE *identifies an ϵ-optimal arm after $\widetilde{\mathcal{O}}(\frac{1}{\epsilon} \cdot \log 1/\delta)$ samples.*

Proposition 1 implies that, to solve the safe best-arm identification problem optimally, more care must be taken in exploring than either standard experiment design induces—we must trade off between $\mathcal{XY}_{\text{diff}}$ and $\mathcal{XY}_{\text{safe}}$ as BESIDE does. We remark briefly on the instance $\mathcal{I}^1$. On this instance we have $\mathcal{X} = \{e_1, e_2\}$ and $\mathcal{Z} = \{z_1, z_2\}$ with $z_1 = [1/4, 1/2]$ and $z_2 = [3/4, 1/2 + \alpha]$. We set $\theta^1_* = [1, 0]$, $\mu^1_* = [0, 1]$, and $\gamma = 1/2 + \alpha/2$. Here z_2 is unsafe while z_1 is safe, so it follows that $z_* = z_1$. As $z_2^\top \theta^1_* > z_1^\top \theta^1_*$, to show $z_2 \neq z_*$, we must show it is unsafe. However, if we solve the design $\mathcal{XY}_{\text{diff}}$, we see that it places nearly all of the mass on the first coordinate. While this would be optimal if both z_1 and z_2 were safe and we simply wished to determine which has a higher value, to show z_2 is unsafe, the optimal strategy places (roughly) the same mass on each coordinate, since each coordinate could contribute to the safety value. This is precisely the allocation BESIDE will play, so it is able to show that z_2 is unsafe much more efficiently than a naive $\mathcal{XY}_{\text{diff}}$ approach.

6.3 Experiments for Safe Best Arm Identification in Linear Bandits

We next present experimental results on BESIDE to demonstrate the advantage of experimental design—especially combining $\mathcal{XY}_{\text{diff}}$ and $\mathcal{XY}_{\text{safe}}$ designs. As there are no existing algorithms that consider safe best-arm identification, as a benchmark we consider the naive adaptive approach BASELINE that first solves the problem of finding the safe arms up to a desired tolerance, and then solves the problem of finding the best (safe) arm among the arms that were found to be safe. We first describe instances on which we test BESIDE. Our experimental details and precise implementation of BESIDE using elimination are described in Section E.6.

Multi-Armed Bandit. We consider a best-arm identification problem in which every arm is safe, but the arm with highest value is very difficult to identify as safe, while the second-best arm can easily be shown safe. We vary the total number of arms and run BESIDE and BASELINE with $\epsilon = 0.5$ and $\delta = 0.1$. From Figure 6.2, we observe that the sample complexity of BESIDE is smaller (up to about two times for 100 arms) than the sample complexity of its baseline.

Linear Response Model. *Random Instance*: We also consider the more general setup where $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$, $\theta \in \mathbb{R}^d$ and $\mu \in \mathbb{R}^d$ are randomly generated from independent Gaussian random variables with mean 0 and variance 1. We set $|\mathcal{X}| = 50$ and vary the size of $|\mathcal{Z}|$. In Figure 6.3, we see again that BESIDE significantly outperforms the baseline.

Hard Instance: We last consider the instance of Proposition 1 and benchmark against the strategy playing only allocations minimizing $\mathcal{XY}_{\text{diff}}$. In Figure 6.4, we see again that BESIDE significantly outperforms this baseline, corroborating the theoretical result of Proposition 1.

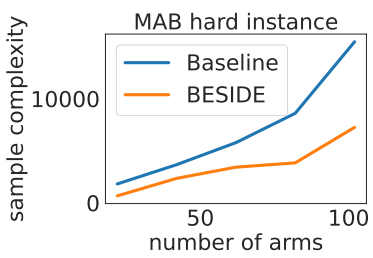


Figure 6.2: Total arm pulls to termination vs. number of arms

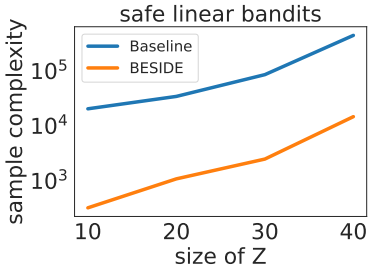


Figure 6.3: Total arm pulls to termination vs. $|\mathcal{Z}|$

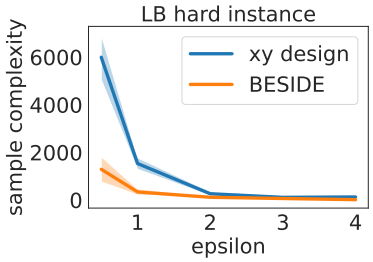


Figure 6.4: Total arm pulls to termination vs. ϵ

6.3.1 Practical Algorithms for Active Classification Under Constraints

Next, we provide an application of the above ideas to pool-based active classification with constraints—namely, adaptive sampling to learn the highest accuracy classifier with a constraint on the false discovery rate (FDR). We first explain how this problem maps to the linear bandit setting. Precisely, let $\mathcal{X}$ be the example space and $\mathcal{Y} = \{0, 1\}$ the label space. Fix a hypothesis class $\mathcal{H}$ such that each $h \in \mathcal{H}$ is a classifier $h : \mathcal{X} \to \mathcal{Y}$. We represent each h with an associated indicator vector $z_h \in \{0, 1\}^{|\mathcal{X}|}$ where $z_h(x) = 1 \iff h(x) = 1$. Similarly, let $\eta \in [0, 1]^{|\mathcal{X}|}$ represent the label distribution, i.e. $\eta(x) = \mathbb{P}(Y = 1|X = x)$. Then the risk of a classifier $R(h) := \mathbb{E}_{x\sim\text{Unif}(\mathcal{X}),Y\sim\textbf{Ber}(\eta(x))}[\mathbf{1}[h(x) \neq Y]] = z_h^\top(2\eta - 1)$ and the FDR is defined as $\text{FDR}(h) := (\mathbf{1} - \eta)^\top z/\mathbf{1}^\top z$. In the case when $\eta \in \{0, 1\}^{|\mathcal{X}|}$, $\text{FDR}(h)$ is the proportion of examples that h incorrectly labels as 1 out of all examples h labels as 1. Our goal is to solve the following constrained best arm identification problem:

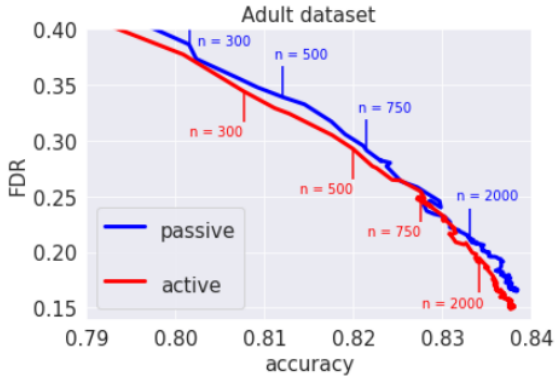
$$\widehat{h} = \min_{h\in\mathcal{H}} R(h) \quad \text{s.t.} \quad \text{FDR}(h) \leq q \iff \min_{h\in\mathcal{H}} z_h^\top \eta \quad \text{s.t.} \quad ((\mathbf{1} - \eta)^\top - q\mathbf{1}^\top)^\top z \leq 0. \tag{6.5}$$

The main challenge in running BESIDE on this problem directly is a potentially high computational cost from computing a design over an extremely large hypothesis class $\mathcal{H}$ (e.g. neural networks of a bounded width). In this section we provide an alternative approach motivated by BESIDE. Algorithm 6.2 follows a similar design as BESIDE and relies on an oracle, CERM, that can solve (6.5), i.e. given a dataset it returns the highest accuracy classifier under an FDR constraint. Such oracles are available in, for example in (Agarwal et al., 2018; Cotter et al., 2018). In each round of Algorithm 6.2 we perform *randomized exploration* by perturbing the labels on our existing dataset with mean zero Gaussian noise, and then training k classifiers $\widehat{h}_i, i \in [k]$, on the resulting datasets. Implicitly, we are making the assumption that the loss function in the training of ERM can handle continuous labels, such as the MLE of logistic regression. As described in (Kveton et al., 2019), randomized exploration emulates sampling from a posterior distribution on our possible set of classifiers. We then use the labels generated from these classifiers to compute safe classifiers $h_i, i \in [k]$. Finally, mimicking the strategy of BESIDE, we compute $\mathcal{XY}_{\text{safe}}$ and $\mathcal{XY}_{\text{diff}}$ designs on these k safe classifiers and repeat (note that the designs computed on Line 5 are equivalent to $\mathcal{XY}_{\text{safe}}$ and $\mathcal{XY}_{\text{diff}}$ in the classification setting).

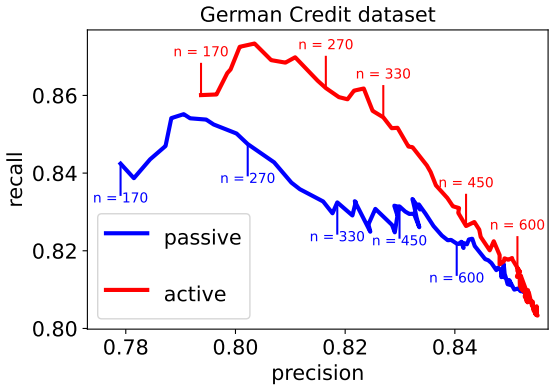
Algorithm 6.2. `Active constrained classification with randomized exploration`

Input: Batch size n, initial (labeled) data $x_1^{(0)}, \ldots, x_n^{(0)}$, number of rounds L, number of classifiers per round k, perturbation variance σ

1: **for** $\ell = 1, \ldots, L$ **do**
2: **for** $i = 1, \ldots, k$ **do**
3: $\widehat{h}_i = \texttt{ERM}(\{(x_t^{(\ell)}, y_t^{(\ell)} + \epsilon_t^{(i)})\}_{t=1}^n)$, where $\{\epsilon_t^{(i)}\}_{1 \le t \le n} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$
4: $h_i = \texttt{CERM}(\{(x, \widehat{h}_i(x))\}_{x \in \mathcal{X}})$
5: Compute designs: $\lambda_{\text{safe}} = \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{1 \le i \le k} \sum_{x \in \mathcal{X}} \frac{\mathbf{1}\{h_i(x) \neq 0\}}{\lambda_x}$, $\lambda_{\text{diff}} = \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{1 \le i \ne j \le k} \sum_{x \in \mathcal{X}} \frac{\mathbf{1}\{h_i(x) \neq h_j(x)\}}{\lambda_x}$
6: Sample $x_1^{(\ell)}, \ldots, x_n^{(\ell)}$ from a uniform mixture of $\lambda_{\text{safe}}, \lambda_{\text{diff}}$
7: Observe corresponding labels $y_1^{(\ell)}, \ldots, y_n^{(\ell)}$
return $\widetilde{h} = \texttt{CERM}(\{(x_t^{(\ell)}, y_t^{(\ell)})\}_{1 \le t \le n, 0 \le \ell \le L})$



(a) FDR vs accuracy for active (Algorithm 6.2) and passive sampling, ticks report number of samples. FDR and accuracy are averaged over 5 trials



(b) TPR vs FDR for active active (Algorithm 6.2) and passive sampling, ticks report number of samples. Precision is $1 - \text{FDR}$, recall is TPR. Precision and recall are averaged over 25 trials

To validate Algorithm 6.2, we experiment against a passive baseline that selects points uniformly at randoms from the pool of examples $\mathcal{X}$, retrains the model using the same Constrained Empirical Risk Minimization oracle (`CERM`) as Algorithm 6.2 on its current samples, and report the accuracy and FDR. We evaluate on two real world datasets and on one synthetic dataset next and provide an additional details on the experiments in Section E.6.

Adult dataset. We evaluate on the adult income data set (Lichman, 2013) (48,842 examples) where the goal is to predict whether someone's income is above \$50k per year. We set the constraint to be FDR < 0.15 and report in Figure 6.5a the accuracy and the FDR obtained when varying the number of labels given to each method (batch size is set to 25 and initial number of queried labels is 50). We observe that for any desired accuracy Algorithm 6.2 allows us to provide a classifier with lower FDR. Also, for any chosen number of total labels—such as $500, 750, 2000$ as reported in Figure 6.5a—the Algorithm 6.2 gives a classifier with higher accuracy and lower FDR. In general we found that the active method needed half the number of samples as the passive sampling to achieve a given FDR. This demonstrates the effectiveness of Algorithm 6.2 to learn simultaneously the objective (risk) and the constraint (FDR), in a similar favorable

way as characterized by our theoretical findings.

We consider the German Credit Dataset originally from the Staflog Project Databases (Keogh et al., 1998). The goal is to predict whether someone's credit is 'bad' or 'good'. We report in Figure 6.5b the recall (TPR) and the precision ($1 - \text{FDR}$) obtained when varying the number of labels given to each method. We observe that for any desired precision Algorithm 6.2 allows us to provide a classifier with higher recall. Also, for any chosen number of total labels—such as $170, 270, 330, 450, 600$ as reported in Figure 6.5b—the Algorithm 6.2 gives a classifier with higher precision and higher recall. As for the Adult dataset we found that the active method needed half the number of samples as the passive sampling to achieve a given precision.

Half circle dataset. We consider a two-dimensional half circle dataset, visualized on Figure 6.6. We report in Figures 6.7 and 6.8 the precision and (respectively) the recall obtained when varying the number of labels given to each method. The confidence intervals are obtained over 25 repetitions. We observe that Algorithm 6.2 allows us to provide a classifier satisfying a given recall or precision in far fewer queries. This is in line with the results of (Jain and Jamieson, 2020) on One Dimensional Thresholds, where the sample complexity of the active strategy is $O(log(n))$ while the sample complexity of the passive strategy is at least of order n.

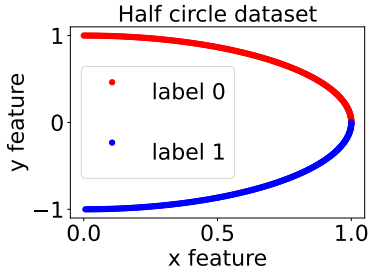


Figure 6.6: Half circle dataset.

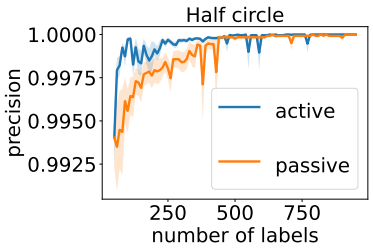


Figure 6.7: Precision

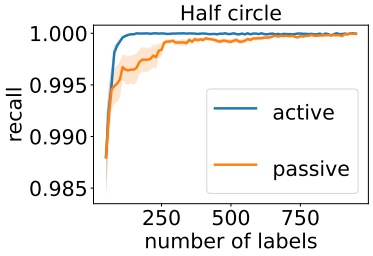


Figure 6.8: Recall

6.4 Related works

Constrained Bandits. A growing body of work seeks to address the question of safe learning in interactive environments. In particular, the majority of such works have considered the problem of regret minimization in linear bandits with linear safety constraints. Here, the goal is to maximize online reward, $x_t^\top \theta_*$, by choosing actions $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$, while ensuring a safety constraint of the form $x_t^\top \mu_* \leq \gamma$ is met at all times (either in expectation or with high probability). A variety of algorithms have been proposed, including UCB-style (Kazerouni et al., 2017; Amani et al., 2019; Pacchiano et al., 2020), and Thompson Sampling (Moradipari et al., 2019; 2020). While these works show that $\sqrt{T}$ regret is attainable, they only provide worst-case bounds (while we obtain instance-dependent bounds) and do not study the pure-exploration best-arm identification problem. To our knowledge, the only work to offer instance-dependent guarantees is (Chen et al., 2022), yet they focus exclusively on the regret setting, and offer a relatively coarse notion of instance-dependence — analogous to $\mathcal{O}(d \cdot \text{poly} \log T/\Delta_{\min})$ bounds in the unconstrained linear bandits setting — in contrast to the more fine-grained notion of instance-dependence we provide.

To our knowledge, only several existing works consider the question of best-arm identification with safety constraints (Sui et al., 2015; 2018; Wang et al., 2022; Lindner et al., 2022). The most related to ours

is (Lindner et al., 2022) which focuses on the easier problem of safe best arm identification with known rewards and unknown constraints. Since the reward is known, the main challenge in the setting of (Lindner et al., 2022) is to learn constraints via G-optimal designs. The key and novel challenge of our framework is to carefully balance between G and $\mathcal{XY}$ designs: naively spending enough budget to either learn the reward model (via a $\mathcal{XY}$ design) or to learn the safety constraints (via a G design) will fail catastrophically (see Example 2.1 and Proposition 1). (Sui et al., 2015; 2018) consider a general constrained optimization setting where the goal of the learner is to minimize some function $f(x)$ over a domain $x \in \mathcal{D}$, while only having access to noisy samples of $f(x)$, $f(x_t) + w_t$, and guaranteeing that a safety constraint $g(x_t) \geq h$ is met for every query point x_t. While they do provide a sample complexity upper bound, they give no lower bound, and, as shown in (Wang et al., 2022), their approach can be very suboptimal. (Wang et al., 2022) considers the setting of best-arm identification in multi-armed bandits. In their setting, at every step t they query a value $a_t \in \mathcal{A}$ for a particular coordinate i_t, and their goal is to identify the coordinate i^* such that $a^*_{i^*}\theta_{i^*} \geq \max_i a_i^*\theta_i$, where a_i^* is the largest value respecting the safety constraint: $a_i^* = \arg\max_{a \in \mathcal{A}} a\theta_i$ s.t. $a\mu_i \leq \gamma$. Similar to (Sui et al., 2015; 2018), they require that the safety constraint $a_t\mu_{i_t} \leq \gamma$ must be met while learning. Though they do show matching upper and lower bounds, and in addition consider a slightly more general setting that allows for nonlinear (but monotonic) response functions, they treat every coordinate as independent, and do not allow for information-sharing between coordinates—the key generalization the linear bandit setting targets. We remark as well that in our setting, unlike these works, we allow the learner to query unsafe points during exploration, and only require that they output a safe decision at termination.

Best-Arm Identification in Linear Bandits. The best-arm identification problem in multi-armed bandits (without safety constraints) is a classical and well-studied problem (Bechhofer, 1958; Paulson, 1964; Even-Dar et al., 2002; Bubeck et al., 2009), and near-optimal algorithms exist (Jamieson et al., 2014; Kaufmann et al., 2016). More recently, there has been a growing interest in understanding the sample complexity of best-arm identification in linear bandits (Soare et al., 2014; Karnin, 2016; Xu et al., 2018; Fiez et al., 2019; Katz-Samuels et al., 2020; Degenne et al., 2020). We highlight in particular the work of (Fiez et al., 2019) which proposes an experiment-design based algorithm, RAGE, that our approach takes inspiration from. While much progress has been made in understanding best-arm identification in linear bandits, to our knowledge, no existing works consider the setting of best-arm identification in linear bandits with safety constraints, the setting of this work.

Active Classification under FDR constraints We finally mention one other related body of work—the problem of actively sampling to find a classifier with high accuracy or recall under precision constraints. Motivated by the experimental design approach of our main algorithm, BESIDE, we provide a heuristic algorithm for this problem with good empirical performance in Section 6.3.1. There is an extensive body of work on active learning (see the survey (Hanneke et al., 2014)) but only recently have works made the connection between best-arm identification for linear bandits and classification (Katz-Samuels et al., 2021; Jain and Jamieson, 2020; Camilleri et al., 2021b). Precision constraints has been less studied in the adaptive context, we only know of (Jain and Jamieson, 2020; Bennett et al., 2017).

6.5 Conclusion

In this work we have shown that it is possible to efficiently find the best *safe* arm in linear bandits with a carefully designed adaptive experiment design-based approach. Our results open up several interesting directions for future work.

Instance Optimality. While BESIDE is worst-case optimal, in Section E.1 we show an instance-dependent lower bound which BESIDE does not, in general, seem to hit. We conjecture that this lower bound may be loose—addressing this discrepancy and showing matching instance-dependent upper and lower bounds is an exciting direction for future work.

Safety During Exploration. Though there are many interesting applications where we may not require safety during exploration (i.e. only querying safe arms), in other cases we may need to ensure safety is met during exploration. Extending our work to this setting is an interesting open problem.

Chapter 7

Fair Active Learning in Low-Data Regimes

7.1 Introduction

As machine learning models proliferate and are used in an ever-increasing number of applications with societal ramifications, it has become increasingly important to have robust methods for developing models that do not perpetuate existing social inequities. Over the last few years, a plethora of works in fair classification have provided a principled toolkit to develop classifiers and quantify their performance under various fairness metrics. These metrics, including equal opportunity and equalized odds, give a natural way to ensure that favorable outcomes such as model performance or predicted positive rates are equalized across different groups for a given protected feature. More precisely, given a distribution ν on $\mathcal{X} \times \mathcal{A} \times \mathcal{Y}$ (where $\mathcal{X}$ is the feature space, $\mathcal{A}$ the protected attribute space and $\mathcal{Y}$ the label space), a hypothesis class $\mathcal{H}$, a fairness metric m_{fair}, a measure of its violation $L_\nu^{m_{\text{fair}}}(h)$, and a fairness violation tolerance α; the goal in fair classification is to return $\arg\min_{h \in \mathcal{H}} \mathbb{E}_{(x,a,y)\sim\nu}[h(x) \neq y]$ subject to $L_\nu^{m_{\text{fair}}}(h) \leq \alpha$.

In practice, as ν is unknown, solving an empirical analog of this constrained classification problem on a training set is a natural approach to learning classifiers that generalize well to a test set, while maintaining fairness guarantees. Indeed, the focus of much of the fairness literature has been to develop optimization methods to solve such a problem (Agarwal et al., 2018; Donini et al., 2018; Cotter et al., 2018). While this is a reasonable approach when a large amount of labeled training data is available, in many applications such large amounts of data are not available, and it can be prohibitively expensive to collect more. In such settings existing approaches may not be able to guarantee accurate classifiers, or may return classifiers that are in fact unfair on the population distribution.

A promising approach to handle such low-data regimes and maximize the effectiveness of small amounts of labeled data is *active learning*. Active learning methods aim to minimize the amount of labeled training data needed by only requesting labels for the most *informative* examples, thereby significantly reducing the label complexity while ensuring similar accuracy of the learned classifier. While active learning methods have been applied to fair classification before, existing works either require large labeled datasets for pretraining, thereby eliminating the primary benefit of active learning, or are unable to satisfy the goal fairness constraint.

In this chapter we aim to overcome these challenges and develop methods for fair active learning which do not require large pretraining datasets—truly operating in the low-data regime—and ensure fairness constraints are met. Our contributions are as follows:

1. We propose a novel approach to fair active learning, FARE, which chooses which points to label by combining a posterior sampling-inspired randomized exploration procedure that aims to improve classifier

accuracy, with a group-dependent sampling procedure to ensure fairness is met. Notably, our approach does not require a large pretraining dataset, and is able to produce accurate and fair classifiers in the very low data regime.

2. We evaluate our proposed method on a variety of standard benchmark datasets from the fairness community, and demonstrate that it yields large label complexity gains over passive approaches while ensuring fairness constraints are met, and also significantly outperforms the existing state-of-the-art approaches for fair active learning.

To the best of our knowledge, our proposed approach is the first active learning procedure able to ensure fairness constraints are reliably met without requiring large amounts of labeled data.

7.2 Related Work

Fairness. Algorithmic fairness has garnered significant interest in recent years (see Barocas et al. (2017); Hort et al. (2022) for recent surveys). Approaches to mitigate fairness disparities can be grouped into three lines of work: pre-processing, in-processing, and post-processing. Pre-processing aims to remove disparate impact by modifying the training data(Kamiran and Calders, 2012), while post-processing modifies already learned classifiers to improve fairness (Hardt et al., 2016). Of particular interest to our work is in-processing for bias mitigation, where the focus is on modifying the learning process to build fair classifiers (Zhang et al., 2018). Most relevant to us within in-processing bias mitigation techniques are works that have approached fairness mitigations in classification as a constrained optimization problem (Agarwal et al., 2018; Donini et al., 2018). Our fairness metrics of interest—equal opportunity and equalized odds—were introduced as operationalizations of fairness concurrently by Hardt et al. (2016); Kleinberg et al. (2016); see also (Kearns et al., 2018).

Active learning. The expense associated with labeling data has emerged as a significant obstacle in the practical implementation of machine learning methods. Motivated by this, there has been growing attention towards the concept of *active* classification, which involves presenting the learner with a set of unlabeled examples, and tasking them with producing a precise hypothesis after querying as few labels as possible (Settles, 2011). Active learning has been studied extensively over the past five decades (see the survey Hanneke et al. (2014)). Most active learning approaches select samples to label based on some notion of uncertainty (e.g., entropy of predictions, margin, disagreement (Cohn et al., 1994; Beygelzimer et al., 2009)). Recent breakthroughs have connected best-arm identification for linear bandits with classification, opening up new possibilities for active learning via *experiment design* (Katz-Samuels et al., 2021; Camilleri et al., 2022; 2021b).

Fair active learning. The problem of fair active classification has been previously considered by recent efforts to reach a classifiers with good "fairness-error" trade-off given a label budget, including Anahideh et al. (2021); Sharaf et al. (2022); Fajri et al. (2022). As we will see experimentally, these works suffer from a variety of shortcomings: for example, poor generalization of their fairness violation, minimal accuracy gains over baseline methods, or limited ability to handle standard group fairness metrics. Furthermore, their objective is somewhat different than ours. While we aim to return a classifier with fairness violation below a desired tolerance (motivated by situations where it is critical to ensure our classifier satisfies a given fairness constraint), these works instead aim to quantify the general tradeoff between fairness and accuracy, without ensuring the returned classifier is below any tolerance. Last, these works assume the existence of

large, pre-existing, labeled datasets: namely for their experiments on the `Adult income` dataset Anahideh et al. (2021); Sharaf et al. (2022); Fajri et al. (2022) assume respectively that $2000, 15000, 3000$ labels are accessible. We will see that the gains from our active learning algorithms are instead visible after collecting 100 labels. Other works, such as Cao and Lan (2022b) focuses on fair active learning for decoupled models and Shen et al. (2022); Cao and Lan (2022a), have focused on the analogous problem of finding classifiers that meet *metric*-fair constraints, while Abernethy et al. (2021); Shekhar et al. (2021); Cai et al. (2022); Branchaud-Charron et al. (2021) have focused on data collection for *min-max* fairness. The nature of min-max fairness does not explicitly constrain the differences in quantities between groups, instead improving the quantity for the worst-off group as much as possible. These, alongside the metric fairness constraints, are significantly different than the group fairness metrics we consider, and as such motivate an entirely different set of methods.

Another related line of work is that of bandits with constraints (Sui et al., 2015; Kazerouni et al., 2017; Pacchiano et al., 2020; Wang et al., 2022; Camilleri et al., 2022). As noted, classification can be modeled as a bandit problem and in some cases bandit algorithms can be applied to active learning for classification. Furthermore, imposing unknown constraints in bandit problems is similar to imposing fairness constraints in classification. To the best of our knowledge, however, existing work on constrained bandits does not consider constraints expressive enough to encode standard fairness metrics such as equalized odds and equal opportunity.

7.3 Preliminaries

In the work presented in this chapter, we focus on a binary classification scenario where each data point consists of three elements (x, a, y). Here, $x \in \mathcal{X} \subset \mathbb{R}^d$ represents a d-dimensional feature vector, $a \in \{0, 1\}$ indicates a binary protected attribute which partitions our data into two *groups*, and $y \in \{0, 1\}$ denotes a label. In the general classification paradigm, we assume that the training set $\mathcal{D} = \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\} \sim \nu \in \triangle_{\mathcal{X}\times\{0,1\}\times\{0,1\}}$ is a set of n examples sampled from a target distribution ν. The objective is to learn from the training set $\mathcal{D}$ a classifier $h : \mathcal{X} \mapsto \{0, 1\}$ among a hypothesis set $\mathcal{H}$ (e.g. linear classifiers or random forests) which has the lowest risk $R_\nu(h)$ possible on the target distribution. Here the risk is defined for any distribution $\nu \in \triangle_{\mathcal{X}\times\{0,1\}\times\{0,1\}}$ as $R_\nu(h) := \mathbb{E}_{(x,a,y)\sim\nu}[\mathbf{1}\{h(x) \neq y\}]$.

7.3.1 Definitions of Fairness

In this work we consider in particular two well-known definitions of fairness: Equal Opportunity—also called True Positive Rate Parity (TPRP)—and Equalized Odds (EO), though our method extends to other notions of fairness as well. We formally define these here.

Definition 4 (Fairness Definitions (EO, TPRP))**.** *Given a tolerance $\alpha \in [0, 1]$ and target distribution ν, a classifier $h \in \mathcal{H}$ satisfies True Positive Rate Parity up to α on ν if*

$$|P_{(x,a,y)\sim\nu}(h(x) = 1|a = 0, y = 1) - P_{(x,a,y)\sim\nu}(h(x) = 1|a = 1, y = 1)| \leq \alpha. \tag{7.1}$$

A classifier satisfies Equalized Odds up to α on a distribution ν if, in addition to satisfying (7.1) *it also satisfies*

$$|P_{(x,a,y)\sim\nu}(h(x) = 1|a = 0, y = 0) - P_{(x,a,y)\sim\nu}(h(x) = 1|a = 1, y = 0)| \leq \alpha. \tag{7.2}$$

If $\alpha = 0$, EO states that the prediction $h(x)$ is conditionally independent of the protected attribute a given the label y. With these definitions of fairness in mind, we also define the fairness violation of a given classifier as the left-hand sides of equations (7.1) and (7.2).

Definition 5 (Fairness violation)**.** *We define the EO (resp. TPRP) violation of classifier h on distribution ν as*

$$L_\nu^{\mathrm{EO}}(h) := \max_{z\in\{0,1\}} |P_{(x,a,y)\sim\nu}(h(x)=1|a=0,y=z) - P_{(x,a,y)\sim\nu}(h(x)=1|a=1,y=z)|,$$
$$L_\nu^{\mathrm{TP}}(h) := |P_{(x,a,y)\sim\nu}(h(x)=1|a=0,y=1) - P_{(x,a,y)\sim\nu}(h(x)=1|a=1,y=1)|.$$

Given some threshold α, a *fair classifier* is a classifier with fairness violation below α.

7.3.2 Problem Statement

Classical machine learning typically deals with the setting where the learner has access to a fixed, labeled dataset, $\mathcal{D}_{\mathrm{tr}}$, and must learn as accurate a classifier as possible from this data. In this work, we are interested in the *active* setting where the goal of the learner is to train on as few labeled data points as possible to obtain a desired accuracy. In particular, in the pool-based active learning setting, the task of fair active classification is the following sequential problem. First, the learner is given an unlabeled training pool of data $\mathcal{D}_{\mathrm{tr}}^{\setminus y} \subseteq \mathcal{X}\times\mathcal{A}$ and some fairness metric $m_{\mathrm{fair}} \in \{EO, TP\}$ with target fairness violation α. At each time $t = 1, 2, \ldots, T$ the agent then chooses any unlabeled point from the pool $(x_t, a_t) \in \mathcal{D}_{\mathrm{tr}}^{\setminus y}$ and requests its label $y_t \in \{0, 1\}$. After requesting T labels, the agent outputs a classifier $h \in \mathcal{H}$. Its performance is evaluated via the two following metrics: error loss $R_\nu(h)$ and fairness violation $L_\nu^{m_{\mathrm{fair}}}(h)$, for ν the population distribution. Note that we assume that the learner may see the true protected attribute before querying the label for a point—see (Awasthi et al., 2020) for a discussion of the case when the protected attribute is noisy.

7.4 Fair Active Learning

In this section, we present our approach to fair active classification, FARE.

7.4.1 Fair Learning with Fixed Datasets

Before considering the active setting, we first consider the question of finding a fair classifier on a fixed dataset. As the general classification paradigm (i.e. classification without fairness constraints) is known to potentially cause disparities when applied to sensitive tasks (Barocas and Selbst, 2016), significant effort has been invested to develop effective algorithms that balance the goal of classification (learn the most accurate classifier) with fairness (learn a classifier with low fairness violation) on static datasets. Given a target distribution ν, a fairness metric denoted $m_{\mathrm{fair}} \in \{EO, TP\}$ and a fairness violation tolerance $\alpha \in [0, 1]$, this fair classification problem can be stated as the following:

$$\underset{h\in\mathcal{H}}{\text{minimize}} \quad R_\nu(h) \quad \text{subject to} \quad L_\nu^{m_{\mathrm{fair}}}(h) \le \alpha. \tag{7.3}$$

In practice, one cannot solve (7.3) directly, as the population, ν, which $R_\nu(h)$ and $L_\nu^{m_{\mathrm{fair}}}(h)$ depend on, is unknown. Instead, we consider empirical estimates of the risk and fairness constraint. As is standard throughout machine learning, we rely on the plug-in estimate of the empirical risk, $\widehat{R}_{\mathcal{D}}(h) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{h(x_i) \neq$

$y_i\}$. Similarly, throughout the fairness literature, a plug-in estimator is typically also used to estimate the fairness violation (Agarwal et al., 2018; Donini et al., 2018; Cotter et al., 2018). As an example, consider the case of estimating TPRP. Let $\mathcal{D} = \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\}$ denote a set of data and recall that the True Positive Rate (TPR) of each group $z \in \{0, 1\}$ can be written as

$$P_{(x,a,y)\sim\nu}(h(x) = 1|a = z, y = 1) = \frac{\mathbb{E}_{(x,a,y)\sim\nu}[\mathbf{1}\{h(x) = 1, y = 1, a = z\}]}{\mathbb{E}_{(x,a,y)\sim\nu}[\mathbf{1}\{y = 1, a = z\}]}. \tag{7.4}$$

A natural approach to empirically estimate the TPRP is then to simply replace the population quantities with the empirical quantities in (7.4) to estimate the TPR for each group, and then compute the absolute value of the difference of these TPRs. This yields the following empirical estimate of the TPRP violation of a classifier h on the data $\mathcal{D}$:

$$\widehat{L}^{\mathrm{TP}}_{\mathcal{D}}(h) := \left| \sum_{i=1}^{n} \frac{\mathbf{1}\{h(x_i) = 1, y_i = 1, a_i = 1\}}{\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = 1\}} - \sum_{i=1}^{n} \frac{\mathbf{1}\{h(x_i) = 1, y_i = 1, a_i = 0\}}{\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = 0\}} \right|. \tag{7.5}$$

We can estimate the *false-positive rate parity* (FPRP), $\widehat{L}^{\mathrm{FP}}_{\mathcal{D}}(h)$, analogously to (7.5) but with $y_i = 1$ replaced by $y_i = 0$, and estimate the EO violation as the maximum of the empirical estimate of the TPRP violation and the empirical estimate of the FPRP violation, $\widehat{L}^{\mathrm{EO}}_{\mathcal{D}}(h) = \max\{\widehat{L}^{\mathrm{TP}}_{\mathcal{D}}(h), \widehat{L}^{\mathrm{FP}}_{\mathcal{D}}(h)\}$.

Empirical fair classification. Equipped with these empirical estimates, we return to the fair classification problem, (7.3). Given a training set $\mathcal{D} = \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\} \sim \nu$ sampled from a distribution $\nu \in \triangle_{\mathcal{X}\times\{0,1\}\times\{0,1\}}$, a fairness metric denoted $m_{\text{fair}} \in \{EO, TP\}$ and fairness tolerance $\alpha \in [0, 1]$, one can use the empirical estimates of the risk and the fairness violation to approximate (7.3) with the following *empirical* fair classification optimization problem:

$$\underset{h\in\mathcal{H}}{\text{minimize}} \quad \widehat{R}_{\mathcal{D}}(h) \quad \text{subject to} \quad \widehat{L}^{m_{\text{fair}}}_{\mathcal{D}}(h) \leq \alpha. \tag{7.6}$$

Note that solving such a problem is a common approach to fair classification, and can be solved efficiently (Donini et al., 2018; Agarwal et al., 2018). This optimization problem will form the starting-point of our proposed approach, and our algorithms will assume access to a solver for it, which we call the empirical fair oracle—`EFO`. In our experiments we take an approach analogous to Agarwal et al. (2018) to solve (7.6).

7.4.2 Estimation Error and Sampling Bias

In this section we address two additional issues that arise in ensuring our returned classifier is fair. First, estimation error in the fairness constraint, and second, bias introduced by sampling data points in a non-uniform fashion.

Conservative fairness estimates. Since $\widehat{L}^{m_{\text{fair}}}_{\mathcal{D}}(h)$ is only an empirical estimate of $L^{m_{\text{fair}}}_{\nu}(h)$, ensuring that $\widehat{L}^{m_{\text{fair}}}_{\mathcal{D}}(h) \leq \alpha$ does not guarantee that $L^{m_{\text{fair}}}_{\nu}(h) \leq \alpha$, our end goal. The following result gives a precise quantification of the deviation between $\widehat{L}^{m_{\text{fair}}}_{\mathcal{D}}(h)$ and $L^{m_{\text{fair}}}_{\nu}(h)$ in the case where $m_{\text{fair}} = EO$.

Proposition 2. *Let the train set be $\mathcal{D} = \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\} \sim \nu$. Then it holds with probability $1 - \delta$ that, with $c_\delta := 8\sqrt{2\log(2/\delta)}$:*

$$|L^{\mathrm{EO}}_{\nu}(h) - \widehat{L}^{\mathrm{EO}}_{\mathcal{D}}(h)| \leq \frac{c_\delta}{\sqrt{n}} \cdot \max_{0\leq j,k\leq 1} \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} + \mathcal{O}\left(\frac{1}{n}\right).$$

Analogous results hold for TPRP. This bound inspires two important aspects of our approach. First, to ensure fairness is met, it suggests setting the tolerance in (7.6) to a conservative value less than α, in particular subtracting a $\mathcal{O}(\frac{1}{\sqrt{n}})$ term off of α. Adjusting α by this margin has been demonstrated in the past to produce fair classifiers (Woodworth et al., 2017; Thomas et al., 2019), and we show in Figure 7.10 that it is also critical in our active setting. Second, Proposition 2 suggests that in order to estimate the fairness, we need to collect samples for *each protected attribute*, since our estimation error scales inversely with the minimum number of samples collected for either protected attribute. This observation is critical in motivating our active sampling procedure, as we outline in the following section.

Sampling bias correction. In the active learning paradigm, at every step the learner samples a data point $(x_t, a_t) \in \mathcal{D}_{\text{tr}}^{\setminus y}$ from some (chosen) distribution, $\nu_t^{\text{tr}} \in \triangle_{\mathcal{D}_{\text{tr}}^{\setminus y}}$, $(x_t, a_t) \sim \nu_t^{\text{tr}}$. For example, the learner may place higher weight on points that are *informative*, increasing the number of samples from around the decision boundary. While this will ultimately improve the learner's ability to classify, the distribution of the sampled dataset no longer matches that of the original training dataset. This will result in the plug-in estimator for the fairness constraint, for example (7.5), to be biased. We correct for this mismatch using importance weights. For the risk, we recall the definition of the well-known IPS estimator (empirical risk re-weighted with importance weights): $\widehat{R}_{\mathcal{D},\nu^{\text{tr}},\nu}(h) := \frac{1}{n}\sum_{i=1}^{n} \frac{\nu_i}{\nu_i^{\text{tr}}} \mathbf{1}\{h(x_i) \neq y_i\}$, for $(x_i, a_i) \sim \nu^{\text{tr}}$ and y_i and associated label, and ν_i the population weight of point i[1] and ν_i^{tr} the probability ν^{tr} samples point i. It is straightforward to see that this is an unbiased estimator of the true risk. We define the estimator for EO with importance weights next.

Definition 6 (Empirical EO violation with importance weights)**.** *Consider a dataset drawn i.i.d from ν^{tr}, $\mathcal{D} := \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\} \sim \nu^{\text{tr}}$. The empirical estimate of the EO violation of a classifier h on the target distribution ν can be evaluated as*

$$\widehat{L}^{\text{EO}}_{\mathcal{D},\nu^{\text{tr}},\nu}(h) := \max_{z\in\{0,1\}} \left| \frac{\sum_{i=1}^{n} \frac{\nu_i}{\nu_i^{\text{tr}}} \mathbf{1}\{h(x_i)=1, y_i=z, a_i=1\}}{\sum_{i=1}^{n} \frac{\nu_i}{\nu_i^{\text{tr}}} \mathbf{1}\{y_i = z, a_i = 1\}} - \frac{\sum_{i=1}^{n} \frac{\nu_i}{\nu_i^{\text{tr}}} \mathbf{1}\{h(x_i)=1, y_i=z, a_i=0\}}{\sum_{i=1}^{n} \frac{\nu_i}{\nu_i^{\text{tr}}} \mathbf{1}\{y_i = z, a_i = 0\}} \right|.$$

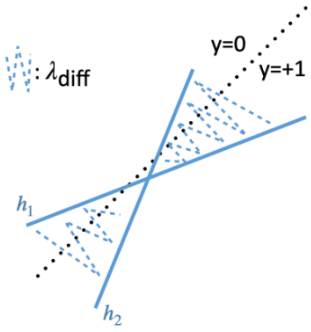
We define the importance-weighted TPRP violation analogously, but only for $z = 1$. While this estimate is not truly unbiased, both the numerator and denominators are unbiased, leading to accurate estimates of the fairness. In the following, when applying our fairness oracle `EFO` in the active setting, we assume it is applied on the importance-weighted fairness and loss estimates.

7.4.3 Fair Active Learning

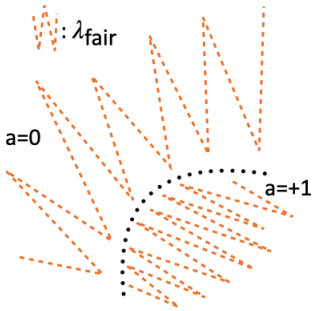
We now provide our algorithm for fair active classification, Algorithm 7.1. Algorithm 7.1 proceeds in rounds. In each round, we choose data points to label by sampling from two distributions: λ_{diff}, which focuses on improving the *accuracy*, and λ_{fair}, which focuses on improving the *fairness estimates*. We describe our choice of each of these distributions below.

Improving accuracy via randomized exploration. In each round of Algorithm 7.1, to determine which points are most likely to improve accuracy, we perform *randomized exploration* by training a set of k fair classifiers $\widehat{h}_i, i \in [k]$, on perturbations of the training data already collected. In particular, to generate these perturbations, while training each classifier $\widehat{h}_i$ we flip the label of each data point with probability σ.

[1] In general this is unknown but, assuming $\mathcal{D}_{\text{tr}}^{\setminus y} \sim \nu$, it suffices to simply set $\nu_i = 1/|\mathcal{D}_{\text{tr}}^{\setminus y}|$



(a) λ_{diff}



(b) λ_{fair}

Figure 7.1: Sampling distributions of FARE when $k = 2$. The oscillating dotted lines are used to represent the support of the sampling distributions (areas where the sampling distribution is non-zero). λ_{diff} places mass on disagreement region of learned classifiers in order to collect points increasing accuracy. λ_{fair} places equal amounts of mass on each group in order to learn fairness value.

Algorithm 7.1. FARE (Fair Active Randomized Exploration)

Input: Batch size n, number of rounds L, classifiers per round k, perturbation rate σ, fairness metric m_{fair}, fairness tolerance α, unlabeled data $\mathcal{D}_{\text{tr}}^{\setminus y}$

Sample $(x_1^{(0)}, a_1^{(0)}), \ldots, (x_n^{(0)}, a_n^{(0)}) \sim \mathbf{U}(\mathcal{D}_{\text{tr}}^{\setminus y})$, request labels for sampled points

$\mathcal{D}_0 \leftarrow \{(x_i^{(0)}, a_i^{(0)}, y_i^{(0)})\}_{i=1}^n$

$\mathcal{D}_{\text{tr}}^{\setminus y} \leftarrow \mathcal{D}_{\text{tr}}^{\setminus y} \setminus \{(x_i^{(0)}, a_i^{(0)})\}_{i=1}^n$

for $\ell = 1, \ldots, L-1$ **do**

// Compute λ_{diff}

  for $i = 1, \ldots, k$ **do**

    $h_i = \texttt{EFO}(\widetilde{\mathcal{D}}_{\ell-1}, \alpha - \frac{1}{\sqrt{n \cdot \ell}})$ where $\widetilde{\mathcal{D}}_{\ell-1}$ generated by flipping each label of $\mathcal{D}_{\ell-1}$ w.p. σ

  Compute λ_{diff} allocation:

$$\lambda_{\text{diff}} \leftarrow \text{argmin}_{\lambda \in \triangle_{\mathcal{D}_{\text{tr}}^{\setminus y}}} \max_{1 \le i \ne j \le k} \sum_{(x,a) \in \mathcal{D}_{\text{tr}}^{\setminus y}} \frac{\mathbf{1}\{h_i(x) \ne h_j(x)\}}{\lambda_x}$$

// Compute λ_{fair}

  $\lambda_{\text{fair}} \leftarrow \frac{1}{2}\mathbf{U}(\{(x,a) \in \mathcal{D}_{\text{tr}}^{\setminus y} \ : \ a = 0\}) + \frac{1}{2}\mathbf{U}(\{(x,a) \in \mathcal{D}_{\text{tr}}^{\setminus y} \ : \ a = 1\})$

// Sample points and update classifier

  Sample $(x_i^{(\ell)}, a_i^{(\ell)}) \sim \frac{1}{2}\lambda_{\text{diff}} + \frac{1}{2}\lambda_{\text{fair}}, i = 1, \ldots, n$

  Observe corresponding labels $y_1^{(\ell)}, \ldots, y_n^{(\ell)}$

  $\mathcal{D}^{\ell} \leftarrow \mathcal{D}^{\ell-1} \cup \{(x_i^{(\ell)}, a_i^{(\ell)}, y_i^{(\ell)})\}_{i=1}^n$

  $\mathcal{D}_{\text{tr}}^{\setminus y} \leftarrow \mathcal{D}_{\text{tr}}^{\setminus y} \setminus \{(x_i^{(\ell)}, a_i^{(\ell)})\}_{i=1}^n$

Return $\widehat{h} = \texttt{EFO}(\mathcal{D}^L, \alpha - \frac{1}{\sqrt{n \cdot L}})$

Given these classifiers, we compute λ_{diff}, which aims to sample unlabeled training points that effectively distinguish between the k classifiers.

As described in a variety of works (Osband et al., 2016; 2018; Russo, 2019; Osband et al., 2019; Kveton et al., 2019; Camilleri et al., 2022), randomized exploration emulates sampling from a posterior distribution

Dataset	Protected Attribute	Dataset Size
`Drug Consumption` (Fehrman et al., 2017)	Gender	1885
`Bank` (Moro et al., 2014)	Education Level	11,162
`German Credit` (Hofmann, 1994)	Gender	1,000
`Adult Income` (Lichman, 2013)	Gender	48,842
`Compas` (Lichman, 2013)	Gender	5,278
`Community and Crime` (Redmond and Baveja, 2002)	Race	1,902

Table 7.1: Benchmark datasets

over the optimal classifier. The sampling distribution λ_{diff} is such that the weights will be large for the points x about which the k classifiers disagree most. Indeed, taking $k = 2$ for illustration, we have $\lambda_{\text{diff}} = \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \sum_{x \in \mathcal{X}} \frac{\mathbf{1}\{h_1(x) \neq h_2(x)\}}{\lambda_x}$. If $h_1(x) = h_2(x)$ then $\frac{\mathbf{1}\{h_1(x) \neq h_2(x)\}}{\lambda_x} = 0$ for any $\lambda_x > 0$. In order to minimize $\sum_{x \in \mathcal{X}} \frac{\mathbf{1}\{h_1(x) \neq h_2(x)\}}{\lambda_x}$, one can set λ_x to be very small at regions of $\mathcal{X}$ where $h_1 = h_2$ and very large at regions of $\mathcal{X}$ where $h_1 \neq h_2$. See Figure 7.1a for an illustration of this. Given this, if we can ensure $\widehat{h}_i, i \in [k]$ disagree on points close to the true decision boundary, then our sampling procedure will ensure that we sample such points, which will enable us to effectively learn an accurate classifier. With this in mind, we hope to create k classifiers that have a decision boundary close to the true decision boundary, yet this is precisely what will be created by posterior sampling, which our procedure mimics. As we will see in the experiments, this sampling strategy effectively collects labels that are informative, increasing accuracy of the learned classifier.

Improving fairness via attribute-dependent exploration. In addition to learning the decision boundary to obtain a classifier with high accuracy, we must also learn the value of the fairness constraint to ensure our final classifier is fair. While λ_{diff} ensures that we sample points close to the decision boundary, it makes no guarantee that we sample points which allow us to accurately estimate our fairness constraint—our choice of λ_{fair} ensures that we do sample enough to accurately estimate the fairness.

As shown in Proposition 2, if we wish to estimate the fairness value of a given classifier, we must ensure that we have collected sufficiently many data points from each group $j \in \{0, 1\}$. λ_{diff} is not guaranteed to sample such points—for example, if we have severe group imbalance, the overall accuracy may be maximized by ignoring the group with many fewer samples, in which case λ_{diff} will focus on only sampling the larger group. To address this, we choose λ_{fair} to sample an equal number of samples from each group, which will ensure that our fairness estimate will converge to the population fairness, as guaranteed by Proposition 2. See Figure 7.1b for an illustration of this. As we demonstrate in Section 7.5.3, this sampling is absolutely critical if our goal is to learn a fair classifier—without this attribute-dependent sampling, naive active learning methods fail to produce fair classifiers.

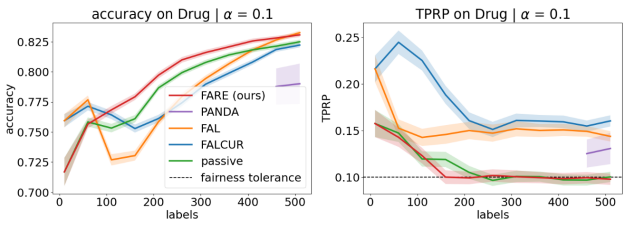


Figure 7.2: Performance on `Drug Consumption`

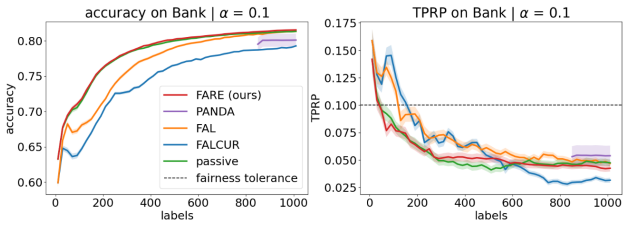


Figure 7.3: Performance on `Bank`

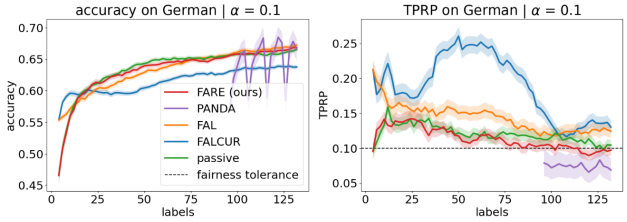


Figure 7.4: Performance on `German Credit`

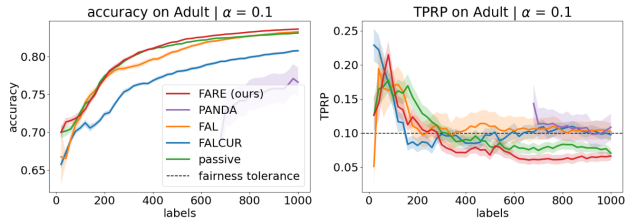


Figure 7.5: Performance on `Adult Income`

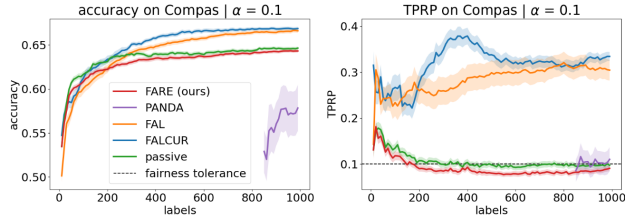


Figure 7.6: Performance on `Compas`

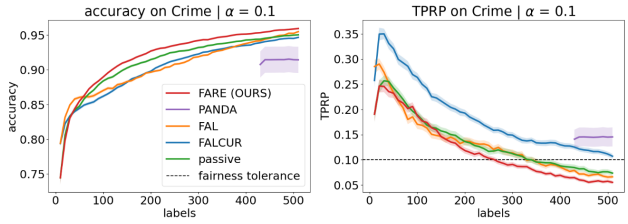


Figure 7.7: Performance on `Community and Crime`

7.5 Experiments

Finally, we demonstrate the effectiveness of FARE experimentally on standard fairness datasets.

Implementation details. For all experiments, we use logistic regression classifiers without regularization and partition the dataset into a $75\%/25\%$ train/test split. We ran a grid-search over the hyperparameters of FARE to set $\sigma = 0.1$ and $k = 10$. We set the fairness tolerance to $\alpha - 1/\sqrt{n}$ to account for estimation error in the fairness constraint. All experiments were run on a Intel Xeon 6226R CPU with 64 cores.

Datasets. In our experiments, we consider six datasets commonly used in the fairness literature, listed in Table 7.1. To ensure consistency, we standardized the data to have a mean of zero and a variance of one.

7.5.1 Baselines Methods

In order to benchmark FARE, we conduct experiments comparing it against state-of-the-art algorithms (Anahideh et al., 2021; Sharaf et al., 2022; Fajri et al., 2022) for fair active learning, and a passive baseline.

1. PANDA (Sharaf et al., 2022): PANDA aims to learn a data selection policy via meta-learning. This algorithm formulates the problem as a bi-level optimization task, where the inner level involves training a classifier with a subset of labeled data, while the outer level focuses on updating the selection policy to strike a balance between fairness and accuracy in the classifier's performance.

	Accuracy (% labeled correctly)					Fairness (TPRP, goal fairness = 0.1)				
	FARE	PANDA	FAL	FALCUR	Passive	FARE	PANDA	FAL	FALCUR	Passive
Drug	**83.1** ± **0.2**	79.0 ± 2.1	83.2 ± 0.2	82.2 ± 0.2	82.5 ± 0.2	0.098 ± 0.006	0.131 ± 0.017	0.144 ± 0.0065	0.160 ± 0.006	0.100 ± 0.005
Bank	**81.5** ± **0.1**	80.1 ± 0.4	**81.3** ± **0.1**	79.2 ± 0.1	**81.3** ± **0.1**	0.042 ± 0.003	0.054 ± 0.009	0.047 ± 0.003	0.032 ± 0.002	0.047 ± 0.001
German	**66.8** ± **0.3**	**66.4** ± **1.4**	67.2 ± 0.4	63.7 ± 0.4	66.6 ± 0.3	0.097 ± 0.007	0.069 ± 0.016	0.124 ± 0.010	0.130 ± 0.010	0.104 ± 0.007
Adult	**83.6** ± **0.0**	76.6 ± 2.0	83.2 ± 0.1	80.8 ± 0.2	83.1 ± 0.0	0.065 ± 0.007	0.109 ± 0.019	0.102 ± 0.013	0.097 ± 0.008	0.068 ± 0.006
Compas	**64.3** ± **0.1**	57.8 ± 1.3	66.6 ± 0.2	66.8 ± 0.2	**64.6** ± **0.2**	0.088 ± 0.006	0.110 ± 0.026	0.304 ± 0.023	0.334 ± 0.009	0.099 ± 0.007
Crime	**95.9** ± **0.1**	91.4 ± 1.7	95.5 ± 0.1	94.7 ± 0.1	95.0 ± 0.1	0.055 ± 0.004	0.145 ± 0.019	0.066 ± 0.004	0.107 ± 0.005	0.074 ± 0.005

Table 7.2: Final accuracy and TPRP values for each method and dataset. Blue indicates fairness threshold met, while red indicates threshold not met. Best accuracy among fair methods is indicated by **bold** font. Confidence intervals are standard errors based on 100 trials.

2. FAL (Anahideh et al., 2021): FAL uses a sampling rule that blends between two selection criteria: one based on uncertainty and another based on assessing fairness, which estimates the potential disparity impact when labeling a specific data point (by calculating the expected disparity across all potential labels). FAL chooses which data points to label in order to strike a balance between model accuracy and equity.

3. FALCUR (Fajri et al., 2022): FALCUR incorporates an acquisition function that assesses the representative score of each sample under consideration. This score is calculated by taking into account two key factors: uncertainty and similarity. By carefully balancing these elements, FALCUR selects samples that contribute to accuracy improvement and ensure that fairness is maintained.

4. Passive + fair oracle: This passive baseline randomly selects points from the pool of examples $\mathcal{D}_{\text{tr}}^{\setminus y}$ and trains the model using the EFO oracle with the same $\alpha - \frac{1}{\sqrt{n}}$ constraint as FARE on its current samples.

Each of these methods with the exception of the passive baseline assumes access to a pretraining dataset. As we are interested in the low-data regime, when we do not have access to a pretraining dataset, we simulate the pretraining dataset by allocating, for each method, some percentage of the label budget to uniform sampling to collect a "pretrain" dataset, and then run the algorithm in standard fashion from there. For each method, we sweep over the size of the pretrain dataset and plot performance for the best one. For all other hyperparameters, we use the values recommended by the original work.

7.5.2 Performance Evaluation

We first consider the case when the fairness constraint is TPRP with $\alpha = 0.1$, and illustrate the accuracy and fairness vs. number of samples for our method and all baselines. For all methods and datasets, with the exception of PANDA, results are averaged over 100 trials—for PANDA results are averaged over only 50 trials, due to its large computational cost. Shaded regions denote one standard error. Note as well that the

	Accuracy (% labeled correctly)					Fairness (TPRP, goal fairness = 0.1)				
	FARE	FARE w/o λ_{fair}	FAL	FALCUR	Passive	FARE	FARE w/o λ_{fair}	FAL	FALCUR	Passive
Synt.	**58.8** **± 0.6**	57.5 ± 0.8	90.0 ± 1.7	89.9 ± 1.3	61.1 ± 0.8	0.095 ± 0.009	0.123 ± 0.016	0.402 ± 0.022	0.303 ± 0.013	0.123 ± 0.013

Table 7.3: Ablation on the role of group-dependent sampling, λ_{fair}, on the synthetically generated dataset. Note that PANDA does not converge on this dataset, so we have omitted it from the table. Confidence intervals are standard errors based on 100 trials.

performance of PANDA starts at a later step since this method requires a large pretrain dataset to perform effectively, and in pretraining does not produce a classifier.

Our results are given in Figures 7.2 to 7.7, and we state the accuracy and fairness values obtained at the final step in Table 7.2. As these results illustrate, FARE consistently outperforms or matches the passive baseline, as well as all existing approaches to fair active classification. We highlight several key features of these results.

First, note that the only methods able to consistently produce classifiers which meet the fairness constraint of $\alpha = 0.1$ are FARE and the passive baselines. While all other methods frequently return classifiers that are unfair, both FARE and the passive baseline return classifiers that, by the final step, are fair on each dataset. We observe that, for very small number of labels, even FARE and the passive baselines produce classifiers which do not meet the fairness constraint—this is to be expected since, for a very small number of samples, it is difficult to estimate the fairness accurately enough to return a fair classifier. We emphasize that, though in some cases the accuracy of FARE is exceeded by baseline approaches, in most situations the baselines do not meet the fairness constraints. Since we are interested in *fair* classification, accuracy values can only be compared in the regime where each classifier is fair.

Second, we highlight the difference in the number of samples required to achieve a given accuracy for FARE as compared to the passive baseline. In particular, on the `Drug`, `Adult`, and `Crime` datasets, FARE requires between 1.4-2x less samples than passive to achieve the final accuracy achieved by passive, while ensuring the fairness constraint is still met. While this gain is not present on every dataset—for `Bank` and `Compas` the performance of FARE and the passive baseline are comparable—these results illustrate that active learning can yield substantial gains over passive approaches for fair classification, while simultaneously ensuring fairness constraints are met.

Fairness Constraints Beyond TPRP. The previously considered results illustrate the performance of each method when the fairness constraint is TPRP. To illustrate the generality of our approach, in Figure 7.8 we also consider the performance of each method when the fairness constraint is equalized odds. As with TPRP, we see that FARE produces a fair classifier while existing approaches fail to, and yields a marked improvement over the passive baseline in terms of accuracy.

Model Selection Beyond Logistic Regression. The aforementioned findings demonstrate how FARE performs when the model is a logistic regression classifier. To showcase the versatility of our method, in Figure 7.9, we compare FARE with passive when the model selection is a decision tree. Similar to logistic regression, we observe that FARE generates a fair classifier and yield a significant accuracy gain compared to the passive baseline.

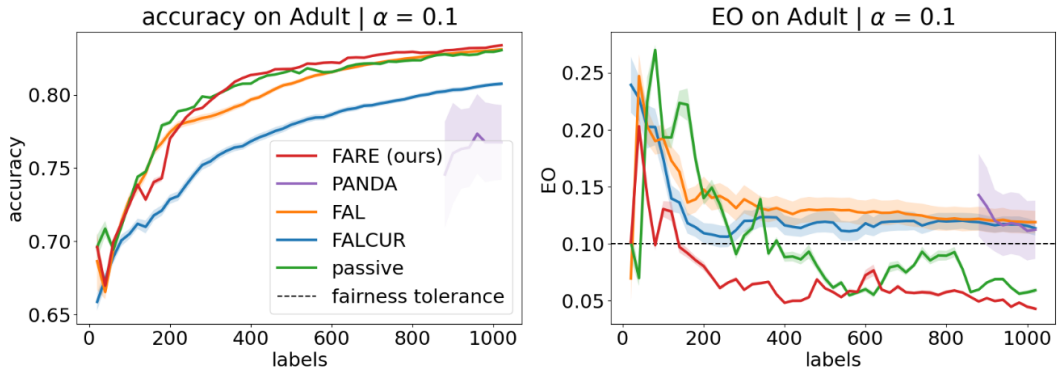


Figure 7.8: Performance on the `Adult Income` dataset for Equalized Odds

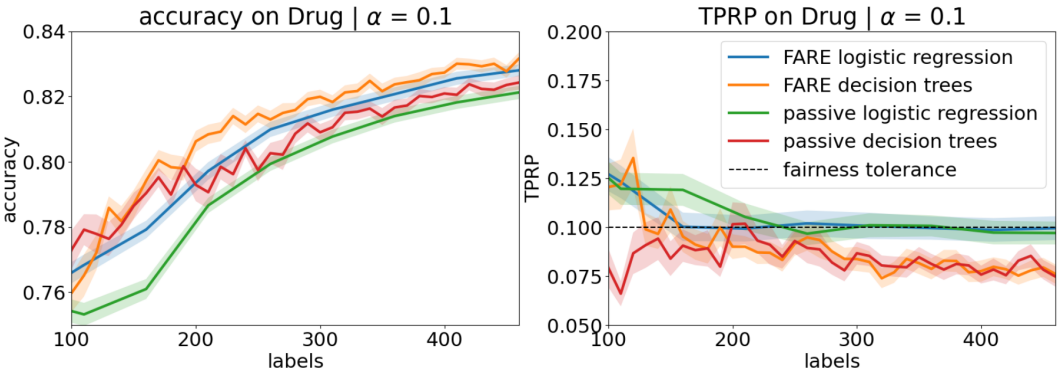


Figure 7.9: Performance on the `Drug` dataset with decision trees

7.5.3 Ablation Experiments

In this section, we illustrate the critical nature of two features of FARE. First, in Figure 7.10, we compare the performance of FARE with the fairness tolerance $\alpha - 1/\sqrt{n}$, with the $1/\sqrt{n}$ term correcting for the estimation error in the fairness constraint, to the performance with the fairness tolerance simply set to α. As shown, with the $1/\sqrt{n}$ correction, the classifier returned by FARE is unfair, while with the correction it is fair. We remark as well that, though the $1/\sqrt{n}$ correction is not precisely what is justified by Proposition 2, this value nonetheless consistently produces fair classifiers.

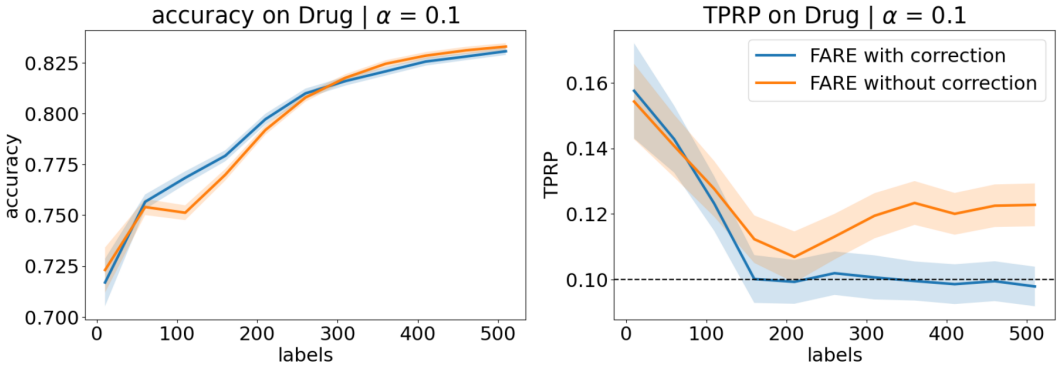


Figure 7.10: Ablation on fairness tolerance correction on `Drug` dataset

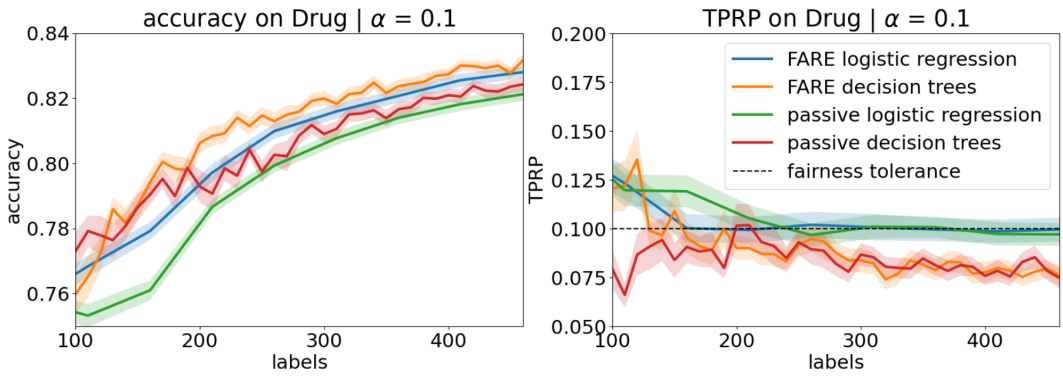


Figure 7.11: Performance on the `Drug Consumption` dataset for logistic regression and decision trees model

Lastly, in Table 7.3, we compare the performance of FARE with and without λ_{fair}, and additionally compare to the performance of the other baselines methods. We evaluate this on a synthetically generated dataset for which there is a large group imbalance—one group has significantly more examples in the dataset than the other. In this setting, if points are not explicitly sampled from the group with the smaller number of examples, virtually all samples will be taken from the larger group, which will cause the fairness estimates to be inaccurate, the resulting classifier unfair. This is illustrated in Table 7.3, where we see that without λ_{fair}, FARE produces an unfair classifier, similar to existing approaches. However, with λ_{fair}, FARE successfully achieves fairness. In conclusion, the inclusion of λ_{fair} in FARE effectively ensures fairness constraints are met, especially when dealing with a significant group imbalance in the dataset.

Generality of FARE. Finally, we delve into the versatility of our proposed method, FARE, by conducting experiments across different machine learning models, specifically comparing its performance on decision trees with that on logistic regression. Our objective is to demonstrate the general applicability of FARE and highlight its consistent effectiveness in guaranteeing fairness fairness across various model architectures.

To assess the generality of FARE, we compare results from experiments conducted on both decision trees and logistic regression models. In Figure 7.11, we showcase the performance of FARE on decision trees compared to logistic regression. Notably, the observed gains in accuracy and the fairness guarantees achieved by FARE on decision trees closely parallel those attained on logistic regression, compared to the passive baseline. This emphasizes the validity of our proposed method across different model types.

7.6 Conclusion

In conclusion, this chapter introduces a novel active learning framework designed to tackle the challenges of bias reduction and accuracy improvement in data-scarce environments critical to machine learning applications. By combining an exploration procedure inspired by posterior sampling with a fair classification subroutine, our proposed approach effectively maximizes accuracy while ensuring fairness constraints in very data-scarce regimes. Through comprehensive evaluations on established real-world benchmark datasets, we demonstrate the efficacy of our framework, highlighting its superiority over state-of-the-art methods. This work contributes to advancing the development of fair models in situations where collecting large labeled datasets is impractical, offering a promising solution for critical applications in machine learning.

Chapter 8

Conclusion

8.1 Impact

This dissertation has had significant impacts across several domains in the field of machine learning, particularly in active learning and bandit problems. Our work has influenced a range of subsequent studies, demonstrating the relevance and applicability of our methods.

- Catoni's Estimator for Bandits (Camilleri et al., 2021a): Our use of Catoni's estimator has extended its application beyond freeing RAGE from its rounding procedure, enabling the development of algorithms for kernel bandits (Camilleri et al., 2021a; Mason et al., 2020), online bandits (Camilleri et al., 2021b), and reinforcement learning (Wagenmaker et al., 2022b).
- Gaussian processes: Our work on Kernel bandits (Camilleri et al., 2021a; Mason et al., 2021) has inspired numerous follow-up studies in Gaussian processes and Bayesian optimization. These studies have improved the handling of infinite action spaces and computational efficiency (Bogunovic and Krause, 2021; Li and Scarlett, 2022; Dai et al., 2024; Zhou and Ji, 2022; Iwazaki et al., 2024; Zenati et al., 2022; Hong et al., 2023; Bogunovic et al., 2022; Shekhar and Javidi, 2022; Shi et al., 2023; Takemori, 2022; Losalka and Scarlett, 2023). Note though that most of these works restrict themselves to regret minimization.
- Constrained Bandits: Our contributions to constrained bandits have prompted further exploration by other researchers (Lindner, 2023; Chen, 2024; Tang et al., 2024; Hutchinson et al., 2024; Lindner et al., 2024; Carlsson et al., 2024).
- Recent and Ongoing Influence: Some of our more recent works are too recent to have been extensively cited yet but are expected to influence future research, such as (Xiong et al., 2024; Camilleri et al., 2023).

8.2 Future Directions

The potential for future research based on this dissertation is vast, with several key areas identified for further exploration:

- Matching Upper and Lower Bounds: Future work should aim to match the upper and lower bounds in several problems addressed by our studies. This is particularly relevant for for Best of Both World

algorithms (Xiong et al., 2024) and Kernel optimization problems (Camilleri et al., 2021a; Mason et al., 2020), the latter being recognized as an open problem (Vakili et al., 2021b).

- G-Design Optimization: Expanding on the optimization of G-designs in infinite spaces for kernel bandits is a promising direction, expanding on the initial findings of this dissertation (Camilleri et al., 2021a; Mason et al., 2020).
- Improving Computational Efficiency: Better optimization formulations are needed to reduce computational costs, particularly in active learning algorithms for streaming settings (Camilleri et al., 2021b).
- Improving Optimization results: Enhancing the guarantees of the optimization subroutine of our online bandits algorithm may drastically improve the computational efficiency of optimal online bandits algorithms (Camilleri et al., 2021b)
- Understanding Randomized Exploration Procedures: A better understanding of randomized exploration procedures will be crucial for advancing experimental design based active learning algorthims (Camilleri et al., 2022; 2023).
- Extending Fair Active Learning to Streaming Settings: Adapting our fair data collection methods (Camilleri et al., 2023) to the streaming data settings will enhance their applicability in real-time decision-making environments.

8.3 Closing words

Throughout our exploration, we encountered various aspects of machine learning and sequential decision making challenges, from high-dimensional experimental design to fairness-aware active learning in low-data regimes. Each chapter presented unique insights and state-of-the-art methodologies, contributing to a comprehensive understanding of the complexities involved in deploying data-driven algorithms responsibly.

Our investigation into high-dimensional experimental design and kernel bandits shed light on the importance of modeling smooth reward functions in Reproducing Kernel Hilbert Spaces (RKHS) and provided crucial insights into novel and nearly optimal algorithms and their performance characteristics, for regret minimization, pure exploration tasks and level set estimation.

Building upon these foundations, our examination of selective sampling for online best-arm identification highlighted the potential of trading-off labeled and unlabeled data in online decision-making processes. We also enabled to ensure safe and efficient model deployment through constrained bandits. Additionally, our exploration of best-arm identification for non-stationary linear bandits underscored the importance of robustness in dynamic environments.

Moreover, our investigation into active learning methodologies integrated with safety and fairness constraints underscored the necessity of aligning model development with ethical considerations and real-world constraints. By integrating safety and fairness constraints into the active learning paradigm, we aimed to enhance the reliability and trustworthiness of machine learning models, ultimately contributing to the advancement of responsible AI.

In conclusion, this thesis represents a concerted effort to navigate the complexities of machine learning deployment, offering insights, methodologies, algorithms, and perspectives aimed at fostering responsible innovation. As we look to the future, it is imperative that we continue to prioritize ethical considerations and societal impacts in the development and deployment of machine learning technologies. By embracing

a holistic approach that encompasses both technical rigor and ethical responsibility, we can strive towards a future where machine learning empowers humanity while upholding our fundamental values.

Bibliography

Y. Abbasi-Yadkori, P. Bartlett, V. Gabillon, A. Malek, and M. Valko. Best of both worlds: Stochastic & adversarial best-arm identification. In *Conference on Learning Theory*, pages 918–949. PMLR, 2018.

J. Abernethy, P. Awasthi, M. Kleindessner, J. Morgenstern, C. Russell, and J. Zhang. Active sampling for min-max fairness, 2021.

A. Agarwal. Selective sampling algorithms for cost-sensitive multiclass prediction. In *International Conference on Machine Learning*, pages 1220–1228. PMLR, 2013.

A. Agarwal, A. Beygelzimer, M. Dudík, J. Langford, and H. Wallach. A reductions approach to fair classification, 2018.

A. E. Alaoui and M. W. Mahoney. Fast randomized kernel methods with statistical guarantees. *arXiv preprint arXiv:1411.0306*, 2014.

Z. Allen-Zhu, Y. Li, A. Singh, and Y. Wang. Near-optimal design of experiments via regret minimization. In *International Conference on Machine Learning*, pages 126–135. PMLR, 2017.

S. Amani, M. Alizadeh, and C. Thrampoulidis. Linear stochastic bandits under safety constraints. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. A. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 9252–9262, 2019. URL `http://papers.nips.cc/paper/9124-linear-stochastic-bandits-under-safety-constraints`.

H. Anahideh, A. Asudeh, and S. Thirumuruganathan. Fair active learning, 2021.

J. Angwin, J. Larson, S. Mattu, and L. Kirchner. Machine bias. In *Ethics of data and analytics*, pages 254–264. Auerbach Publications, 2022.

A. Atkinson, A. Donev, and R. Tobias. *Optimum experimental designs, with SAS*, volume 34. Oxford University Press, 2007.

J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. pages 41–53, 11 2010.

P. Auer and C.-K. Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 116–120. PMLR, 2016.

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 05 2002a. doi: 10.1023/A:1013689704352.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002b.

P. Auer, P. Gajane, and R. Ortner. Adaptively tracking the best bandit arm with an unknown number of distribution changes. In *Conference on Learning Theory*, pages 138–158. PMLR, 2019.

P. Awasthi, M. Kleindessner, and J. Morgenstern. Equalized odds postprocessing under imperfect group information. In *International conference on artificial intelligence and statistics*, pages 1770–1780. PMLR, 2020.

M. J. Azizi, B. Kveton, and M. Ghavamzadeh. Fixed-budget best-arm identification in structured bandits. *arXiv preprint arXiv:2106.04763*, 2021.

D. Azzimonti, D. Ginsbourger, C. Chevalier, J. Bect, and Y. Richet. Adaptive design of experiments for conservative estimation of excursion sets. *Technometrics*, 63(1):13–26, 2021.

F. Bach. On the equivalence between kernel quadrature rules and random feature expansions, 2015.

M.-F. Balcan, A. Beygelzimer, and J. Langford. Agnostic active learning. In *Proceedings of the 23rd international conference on Machine learning*, pages 65–72, 2006.

M.-F. Balcan, A. Z. Broder, and T. Zhang. Margin based active learning. In *Annual Conference Computational Learning Theory*, 2007.

S. Barocas and A. D. Selbst. Big data's disparate impact. *California law review*, pages 671–732, 2016.

S. Barocas, M. Hardt, and A. Narayanan. Fairness in machine learning. *Nips tutorial*, 1:2017, 2017.

A. Barrainkua, P. Gordaliza, J. A. Lozano, and N. Quadrianto. Uncertainty in fairness assessment: Maintaining stable conclusions despite fluctuations. *arXiv preprint arXiv:2302.01079*, 2023.

P. L. Bartlett and S. Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

R. E. Bechhofer. A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. *Biometrics*, 14:408, 1958.

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.

P. N. Bennett, D. M. Chickering, C. Meek, and X. Zhu. Algorithms for active classifier selection: Maximizing recall with precision constraints. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 2017.

R. Berk, H. Heidari, S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth. A convex framework for fair regression, 2017.

D. P. Bertsekas. *Convex optimization theory*. Athena Scientific Belmont, 2009.

A. Beygelzimer, S. Dasgupta, and J. Langford. Importance weighted active learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 49–56, 2009.

A. Beygelzimer, J. Langford, L. Li, L. Reyzin, and R. Schapire. Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 19–26. JMLR Workshop and Conference Proceedings, 2011.

I. Bogunovic. Robust adaptive decision making: Bayesian optimization and beyond. Technical report, EPFL, 2019.

I. Bogunovic and A. Krause. Misspecified gaussian process bandit optimization. *Advances in Neural Information Processing Systems*, 34:3004–3015, 2021.

I. Bogunovic, J. Scarlett, A. Krause, and V. Cevher. Truncated variance reduction: A unified approach to bayesian optimization and level-set estimation. *arXiv preprint arXiv:1610.07379*, 2016.

I. Bogunovic, A. Krause, and J. Scarlett. Corruption-tolerant gaussian process bandit optimization, 2020.

I. Bogunovic, Z. Li, A. Krause, and J. Scarlett. A robust phased elimination algorithm for corruption-tolerant gaussian process bandits. *Advances in Neural Information Processing Systems*, 35:23951–23964, 2022.

D. Bouneffouf, A. Bouzeghoub, and A. L. Gançarski. A contextual-bandit algorithm for mobile context-aware recommender system. In *Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III 19*, pages 324–331. Springer, 2012.

F. Branchaud-Charron, P. Atighehchian, P. Rodríguez, G. Abuhamad, and A. Lacoste. Can active learning preemptively mitigate fairness issues?, 2021.

B. Bryan, J. Schneider, R. Nichol, C. J. Miller, C. R. Genovese, and L. Wasserman. Active learning for identifying function threshold boundaries. In *NIPS*, pages 163–170. Citeseer, 2005.

S. Bubeck and A. Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 42–1. JMLR Workshop and Conference Proceedings, 2012.

S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory: 20th International Conference, ALT 2009, Porto, Portugal, October 3-5, 2009. Proceedings 20*, pages 23–37. Springer, 2009.

J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

W. Cai, R. Encarnacion, B. Chern, S. Corbett-Davies, M. Bogen, S. Bergman, and S. Goel. Adaptive sampling strategies to construct equitable training datasets. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1467–1478, 2022.

X. Cai and J. Scarlett. On lower bounds for standard and robust gaussian process bandit optimization. In *International Conference on Machine Learning*, pages 1216–1226. PMLR, 2021.

R. Camilleri, K. Jamieson, and J. Katz-Samuels. High-dimensional experimental design and kernel bandits. In *International Conference on Machine Learning*, pages 1227–1237. PMLR, 2021a.

R. Camilleri, Z. Xiong, M. Fazel, L. Jain, and K. Jamieson. Selective sampling for online best-arm identification. *Advances in Neural Information Processing Systems*, 34:11071–11082, 2021b.

R. Camilleri, A. Wagenmaker, J. Morgenstern, L. Jain, and K. Jamieson. Active learning with safety constraints. *Advances in Neural Information Processing Systems*, 35:33201–33214, 2022.

R. Camilleri, A. Wagenmaker, J. Morgenstern, L. Jain, and K. Jamieson. Fair active learning in low-data regimes. *arXiv preprint arXiv:2312.08559*, 2023.

Y. Cao and C. Lan. Active approximately metric-fair learning. In *Uncertainty in Artificial Intelligence*, pages 275–285. PMLR, 2022a.

Y. Cao and C. Lan. Fairness-aware active learning for decoupled model. *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9, 2022b. URL `https://api.semanticscholar.org/CorpusID:252625196`.

E. Carlsson, D. Basu, F. Johansson, and D. Dubhashi. Pure exploration in bandits with linear constraints. In *International Conference on Artificial Intelligence and Statistics*, pages 334–342. PMLR, 2024.

R. M. Castro and R. D. Nowak. Minimax bounds for active learning. *IEEE Transactions on Information Theory*, 54:2339–2353, 2007.

O. Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1148–1185, 2012.

N. Cesa-Bianchi, C. Gentile, and F. Orabona. Robust bounds for classification via selective sampling. In *Proceedings of the 26th annual international conference on machine learning*, pages 121–128, 2009.

K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

T. Chen. *Constrained learning in the bandit setting: doubly optimistic strategies and fast rates*. PhD thesis, 2024.

T. Chen, A. Gangrade, and V. Saligrama. A doubly optimistic strategy for safe linear bandits, 2022. URL `https://arxiv.org/abs/2209.13694`.

Y. Chen, C.-W. Lee, H. Luo, and C.-Y. Wei. A new algorithm for non-stationary contextual bandits: Efficient, optimal and parameter-free. In *Conference on Learning Theory*, pages 696–726. PMLR, 2019.

Y. Chen, H. Luo, T. Ma, and C. Zhang. Active online learning with hidden shifting domains. In *International Conference on Artificial Intelligence and Statistics*, pages 2053–2061. PMLR, 2021.

C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Y. Richet. Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465, 2014.

S. R. Chowdhury and A. Gopalan. On kernelized multi-armed bandits, 2017.

D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Machine learning*, 15: 201–221, 1994.

S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pages 797–806, 2017.

A. Cotter, M. Gupta, H. Jiang, N. Srebro, K. Sridharan, S. Wang, B. Woodworth, and S. You. Training well-generalizing classifiers for fairness metrics and other data-dependent constraints, 2018.

Z. Dai, G. K. R. Lau, A. Verma, Y. Shu, B. K. H. Low, and P. Jaillet. Quantum bayesian optimization. *Advances in Neural Information Processing Systems*, 36, 2024.

S. Dasgupta. Coarse sample complexity bounds for active learning. In *NIPS*, 2005.

S. Dasgupta, D. J. Hsu, and C. Monteleoni. A general agnostic active learning algorithm. *Advances in neural information processing systems*, 2008.

S. Dasgupta, A. T. Kalai, and C. Monteleoni. Analysis of perceptron-based active learning. In *Annual Conference Computational Learning Theory*, 2009.

R. Degenne, P. Ménard, X. Shang, and M. Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pages 2432–2442. PMLR, 2020.

O. Dekel, C. Gentile, and K. Sridharan. Selective sampling and active learning from single and multiple teachers. *The Journal of Machine Learning Research*, 13(1):2655–2697, 2012.

M. Derezinski, F. Liang, and M. Mahoney. Bayesian experimental design using regularized determinantal point processes. In *International Conference on Artificial Intelligence and Statistics*, pages 3197–3207. PMLR, 2020.

T. Desautels, A. Krause, and J. Burdick. Parallelizing exploration-exploitation tradeoffs with gaussian process bandit optimization, 2012.

J. Dong and L. F. Yang. Does sparsity help in learning misspecified linear bandits?, 2023.

M. Donini, L. Oneto, S. Ben-David, J. S. Shawe-Taylor, and M. Pontil. Empirical risk minimization under fairness constraints. *Advances in neural information processing systems*, 31, 2018.

M. Donini, L. Oneto, S. Ben-David, J. Shawe-Taylor, and M. Pontil. Empirical risk minimization under fairness constraints, 2020.

S. S. Du, S. M. Kakade, R. Wang, and L. F. Yang. Is a good representation sufficient for sample efficient reinforcement learning?, 2020.

R. Eghbali, J. Saunderson, and M. Fazel. Competitive online algorithms for resource allocation over the positive semidefinite cone. *Mathematical Programming*, 170(1):267–292, 2018.

A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

E. Even-Dar, S. Mannor, and Y. Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *COLT*, volume 2, pages 255–270. Springer, 2002.

R. Fajri, A. Saxena, Y. Pei, and M. Pechenizkiy. Fal-cur: Fair active learning using uncertainty and representativeness on fair clustering, 2022.

E. Fehrman, A. K. Muhammad, E. M. Mirkes, V. Egan, and A. N. Gorban. The five factor model of personality and evaluation of drug consumption risk, 2017.

T. Fiez, L. Jain, K. Jamieson, and L. Ratliff. Sequential experimental design for transductive linear bandits. *NeurIPS*, 2019.

P. I. Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

D. A. Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

M. Gao, J. Huang, X. Huang, S. Zhang, and D. N. Metaxas. Simplified labeling process for medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part II 15*, pages 387–394. Springer, 2012.

A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory: 22nd International Conference, ALT 2011, Espoo, Finland, October 5-7, 2011. Proceedings 22*, pages 174–188. Springer, 2011.

A. Ghosh, S. R. Chowdhury, and A. Gopalan. Misspecified linear bandits, 2017.

A. Gotovos. Active learning for level set estimation. Master's thesis, Eidgenössische Technische Hochschule Zürich, Department of Computer Science,, 2013.

S. Grünewälder, J.-Y. Audibert, M. Opper, and J. Shawe-Taylor. Regret bounds for gaussian process bandit problems. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 273–280. JMLR Workshop and Conference Proceedings, 2010.

K. Guo, R. Cao, X. Kui, J. Ma, J. Kang, and T. Chi. Lcc: towards efficient label completion and correction for supervised medical image learning in smart diagnosis. *Journal of Network and Computer Applications*, 133:51–59, 2019.

A. Gupta, T. Koren, and K. Talwar. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pages 1562–1578. PMLR, 2019.

A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah, K. M. Al-Anazi, et al. An updated review of computer-aided drug design and its application to covid-19. *BioMed research international*, 2021, 2021.

H. Ha, S. Gupta, S. Rana, and S. Venkatesh. High dimensional level set estimation with bayesian neural network. *arXiv preprint arXiv:2012.09973*, 2020.

S. Hanneke and L. Yang. Minimax analysis of active learning, 2014.

S. Hanneke and L. Yang. Toward a general theory of online selective sampling: Trading off mistakes and queries. In *International Conference on Artificial Intelligence and Statistics*, pages 3997–4005. PMLR, 2021.

S. Hanneke et al. Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, 7(2-3):131–309, 2014.

A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, et al. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*, 2014.

M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning, 2016.

E. Hazan and S. Kale. Projection-free online learning. *arXiv preprint arXiv:1206.4657*, 2012.

D. N. Hill, H. Nassif, Y. Liu, A. Iyer, and S. Vishwanathan. An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1813–1821, 2017.

H. Hofmann. Statlog (german credit data) data set. *UCI Repository of Machine Learning Databases*, 53, 1994.

K. Hong, Y. Li, and A. Tewari. An optimization-based algorithm for non-stationary kernel bandits without prior knowledge. In *International Conference on Artificial Intelligence and Statistics*, pages 3048–3085. PMLR, 2023.

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

M. Hort, Z. Chen, J. M. Zhang, F. Sarro, and M. Harman. Bias mitigation for machine learning classifiers: A comprehensive survey. *arXiv preprint arXiv:2207.07068*, 2022.

W. House. Big data: Seizing opportunities and preserving values: In-terim progress report (feb. 2015).

T.-K. Huang, A. Agarwal, D. J. Hsu, J. Langford, and R. E. Schapire. Efficient and parsimonious agnostic active learning. *arXiv preprint arXiv:1506.08669*, 2015.

S. Hutchinson, B. Turan, and M. Alizadeh. Directional optimism for safe linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 658–666. PMLR, 2024.

S. Iwazaki, Y. Inatsu, and I. Takeuchi. Bayesian experimental design for finding reliable level set under input uncertainty, 2019.

S. Iwazaki, S. Takeno, T. Tanabe, and M. Irie. Failure-aware gaussian process optimization with regret bounds. *Advances in Neural Information Processing Systems*, 36, 2024.

L. Jain and K. Jamieson. A new perspective on pool-based active classification and false-discovery control, 2020.

K. Jamieson and L. Jain. Interactive machine learning. 2022.

K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439. PMLR, 2014.

K. G. Jamieson and L. Jain. A bandit approach to sequential experimental design with false discovery control. *Advances in Neural Information Processing Systems*, 31:3660–3670, 2018.

Y. Jedra and A. Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020.

D. Ji, P. Smyth, and M. Steyvers. Can i trust my fairness metric? assessing fairness with unlabeled data and bayesian inference. *Advances in Neural Information Processing Systems*, 33:18600–18612, 2020.

M. Joseph, M. Kearns, J. Morgenstern, and A. Roth. Fairness in learning: Classic and contextual bandits, 2016.

M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth. Meritocratic fairness for infinite and contextual bandits. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596 (7873):583–589, 2021.

K.-S. Jun, L. Jain, B. Mason, and H. Nassif. Improved confidence bounds for the linear logistic model and applications to linear bandits. *arXiv preprint arXiv:2011.11222*, 2020.

N. Kallus and A. Zhou. The fairness of risk scores beyond classification: Bipartite ranking and the xauc metric, 2019.

F. Kamiran and T. Calders. Data preprocessing techniques for classification without discrimination. *Knowledge and information systems*, 33(1):1–33, 2012.

Z. S. Karnin. Verification based solution for structured mab problems. *Advances in Neural Information Processing Systems*, 29, 2016.

Z. S. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, 2013.

Y. Kassahun, B. Yu, A. T. Tibebu, D. Stoyanov, S. Giannarou, J. H. Metzen, and E. Vander Poorten. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *International journal of computer assisted radiology and surgery*, 11:553–568, 2016.

J. Katz-Samuels, L. Jain, Z. Karnin, and K. Jamieson. An empirical process approach to the union bound: Practical algorithms for combinatorial and linear bandits, 2020.

J. Katz-Samuels, J. Zhang, L. Jain, and K. Jamieson. Improved algorithms for agnostic pool-based active classification, 2021.

E. Kaufmann, O. Cappé, and A. Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

A. Kazerouni, M. Ghavamzadeh, Y. Abbasi Yadkori, and B. Van Roy. Conservative contextual linear bandits. *Advances in Neural Information Processing Systems*, 30, 2017.

M. Kearns, S. Neel, A. Roth, and Z. S. Wu. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *International conference on machine learning*, pages 2564–2572. PMLR, 2018.

E. Keogh, C. Blake, and C. J. Merz. Uci repository of machine learning databases,, 1998. URL `http://archive.ics.uci.edu/ml`.

A. Keshavarzi Arshadi, J. Webb, M. Salem, E. Cruz, S. Calad-Thomson, N. Ghadirian, J. Collins, E. Diez-Cecilia, B. Kelly, H. Goodarzi, et al. Artificial intelligence for covid-19 drug discovery and vaccine development. *Frontiers in Artificial Intelligence*, 3:65, 2020.

B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. A. Sallab, S. Yogamani, and P. Pérez. Deep reinforcement learning for autonomous driving: A survey, 2021.

J. Kleinberg, S. Mullainathan, and M. Raghavan. Inherent trade-offs in the fair determination of risk scores, 2016.

R. Kohavi and R. Longbotham. Unexpected results in online controlled experiments. *ACM SIGKDD Explorations Newsletter*, 12(2):31–35, 2011.

B. Kveton, M. Zaheer, C. Szepesvari, L. Li, M. Ghavamzadeh, and C. Boutilier. Randomized exploration in generalized linear bandits, 2019. URL `https://arxiv.org/abs/1906.08947`.

T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

T. Lattimore, C. Szepesvari, and G. Weisz. Learning with good feature representations in bandits and in rl with a generative model, 2020.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

C.-W. Lee, H. Luo, C.-Y. Wei, M. Zhang, and X. Zhang. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously, 2021.

Z. Li and J. Scarlett. Gaussian process bandit optimization with few batches. In *International Conference on Artificial Intelligence and Statistics*, pages 92–107. PMLR, 2022.

M. Lichman. Uci machine learning repository,, 2013. URL `http://archive.ics.uci.edu/ml`.

D. Lindner. *Algorithmic Foundations for Safe and Efficient Reinforcement Learning from Human Feedback*. PhD thesis, ETH Zurich, 2023.

D. Lindner, S. Tschiatschek, K. Hofmann, and A. Krause. Interactively learning preference constraints in linear bandits, 2022. URL `https://arxiv.org/abs/2206.05255`.

D. Lindner, X. Chen, S. Tschiatschek, K. Hofmann, and A. Krause. Learning safety constraints from demonstrations with unknown rewards. *arXiv preprint arXiv:2305.16147*, 2023.

D. Lindner, X. Chen, S. Tschiatschek, K. Hofmann, and A. Krause. Learning safety constraints from demonstrations with unknown rewards. In *International Conference on Artificial Intelligence and Statistics*, pages 2386–2394. PMLR, 2024.

P.-r. Liu, L. Lu, J.-y. Zhang, T.-t. Huo, S.-x. Liu, and Z.-w. Ye. Application of artificial intelligence in medicine: an overview. *Current Medical Science*, 41(6):1105–1115, 2021.

A. Locatelli, M. Gutzeit, and A. Carpentier. An optimal algorithm for the thresholding bandit problem. In *International Conference on Machine Learning*, pages 1690–1698. PMLR, 2016.

A. Losalka and J. Scarlett. Benefits of monotonicity in safe exploration with gaussian processes. In *Uncertainty in Artificial Intelligence*, pages 1304–1314. PMLR, 2023.

G. Lugosi and S. Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

K. Lum and W. Isaac. To predict and serve? *Significance*, 13(5):14–19, 2016.

K. Lum, Y. Zhang, and A. Bower. De-biasing "bias" measurement. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 379–389, 2022.

B. Mason, L. Jain, A. Tripathy, and R. Nowak. Finding all ϵ-good arms in stochastic bandits. *Advances in Neural Information Processing Systems*, 33, 2020.

B. Mason, R. Camilleri, S. Mukherjee, K. Jamieson, R. Nowak, and L. Jain. Nearly optimal algorithms for level set estimation, 2021.

A. C. Miller, L. A. Gatys, J. Futoma, and E. Fox. Model-based metrics: Sample-efficient estimates of predictive model subpopulation performance. In *Machine Learning for Healthcare Conference*, pages 308–336. PMLR, 2021.

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. MIT press, 2018.

A. Moradipari, S. Amani, M. Alizadeh, and C. Thrampoulidis. Safe linear thompson sampling with side information, 2019. URL https://arxiv.org/abs/1911.02156.

A. Moradipari, C. Thrampoulidis, and M. Alizadeh. Stage-wise conservative linear bandits. 2020. doi: 10.48550/ARXIV.2010.00081. URL https://arxiv.org/abs/2010.00081.

S. Moro, P. Cortez, and P. Rita. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62:22–31, 2014.

V. D. Mouchlis, A. Afantitis, A. Serra, M. Fratello, A. G. Papadiamantis, V. Aidinis, I. Lynch, D. Greco, and G. Melagraki. Advances in de novo drug design: from conventional to machine learning methods. *International journal of molecular sciences*, 22(4):1676, 2021.

S. O. Mussmann and S. Dasgupta. Constants matter: The performance gains of active learning. In *International Conference on Machine Learning*, pages 16123–16173. PMLR, 2022.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Stochastic approximation approach to stochastic programming. In *SIAM J. Optim*. Citeseer.

A. Nikolov, M. Singh, and U. T. Tantipongpipat. Proportional volume sampling and approximation algorithms for a-optimal design. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1369–1386. SIAM, 2019.

Optimizely. Stats accelerator – acceleration under time-varying signals. https://support.optimizely.com/hc/en-us/articles/5326213705101-Stats-Accelerator-Acceleration-Under-Time-Varying-Signals, May 2023.

F. Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

I. Osband, B. Van Roy, and Z. Wen. Generalization and exploration via randomized value functions. In *International Conference on Machine Learning*, pages 2377–2386. PMLR, 2016.

I. Osband, J. Aslanides, and A. Cassirer. Randomized prior functions for deep reinforcement learning. *Advances in Neural Information Processing Systems*, 31, 2018.

I. Osband, B. V. Roy, D. Russo, and Z. Wen. Deep exploration via randomized value functions, 2019.

A. Pacchiano, M. Ghavamzadeh, P. Bartlett, and H. Jiang. Stochastic bandits with linear constraints, 2020. URL https://arxiv.org/abs/2006.10185.

E. Paulson. A sequential procedure for selecting the population with the largest mean from k normal populations. *The Annals of Mathematical Statistics*, pages 174–180, 1964.

F. Piccialli, V. S. Di Cola, F. Giampaolo, and S. Cuomo. The role of artificial intelligence in fighting the covid-19 pandemic. *Information Systems Frontiers*, 23(6):1467–1497, 2021.

G. Pleiss, M. Raghavan, F. Wu, J. Kleinberg, and K. Q. Weinberger. On fairness and calibration, 2017.

F. Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

C. Qin and D. Russo. Adaptivity and confounding in multi-armed bandit experiments. *arXiv preprint arXiv:2202.09036*, 2022.

P. Rai, A. Saha, H. Daumé III, and S. Venkatasubramanian. Domain adaptation meets active learning. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, pages 27–32, 2010.

M. Redmond and A. Baveja. A data-driven software tool for enabling cooperative information sharing among police departments. *European Journal of Operational Research*, 141(3):660–678, 2002.

G. Rizk, I. Colin, A. Thomas, and M. Draief. Refined bounds for randomized experimental design. *arXiv preprint arXiv:2012.15726*, 2020.

H. E. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527–535, 1952.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

D. Russo. Worst-case regret bounds for exploration via randomized value functions. *Advances in Neural Information Processing Systems*, 32, 2019.

C. Réda, A. Tirinzoni, and R. Degenne. Dealing with misspecification in fixed-confidence linear top-m identification, 2021.

A. Saha, P. Rai, H. Daumé, S. Venkatasubramanian, and S. L. DuVall. Active supervised domain adaptation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 97–112. Springer, 2011.

J. Scarlett, I. Bogunovic, and V. Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. In *Conference on Learning Theory*, pages 1723–1742. PMLR, 2017.

E. M. Schwartz, E. T. Bradlow, and P. S. Fader. Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4):500–522, 2017.

Y. Seldin and G. Lugosi. An improved parametrization and analysis of the exp3++ algorithm for stochastic and adversarial bandits. In *Conference on Learning Theory*, pages 1743–1759. PMLR, 2017.

Y. Seldin and A. Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning*, pages 1287–1295. PMLR, 2014.

B. Settles. From theories to queries: Active learning in practice. In *Active learning and experimental design workshop in conjunction with AISTATS 2010*, pages 1–18. JMLR Workshop and Conference Proceedings, 2011.

A. Sharaf, H. Daume III, and R. Ni. Promoting fairness in learned models by learning to active learn under parity constraints. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 2149–2156, 2022.

S. Shekhar and T. Javidi. Multiscale gaussian process level set estimation. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3283–3291. PMLR, 2019.

S. Shekhar and T. Javidi. Instance dependent regret analysis of kernelized bandits. In *International Conference on Machine Learning*, pages 19747–19772. PMLR, 2022.

S. Shekhar, G. Fields, M. Ghavamzadeh, and T. Javidi. Adaptive sampling for minimax fair classification. *Advances in Neural Information Processing Systems*, 34:24535–24544, 2021.

J. Shen, N. Cui, and J. Wang. Metric-fair active learning. In *International Conference on Machine Learning*, pages 19809–19826. PMLR, 2022.

Z. Shi, J. Tan, and F. Li. A bayesian approach for bandit online optimization with switching cost. In *Uncertainty in Artificial Intelligence*, pages 1953–1963. PMLR, 2023.

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

M. Simchowitz, K. Jamieson, and B. Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834. PMLR, 2017.

K. Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85, 1918.

M. Soare, A. Lazaric, and R. Munos. Best-arm identification in linear bandits. *arXiv preprint arXiv:1409.6110*, 2014.

N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Y. Sui, A. Gotovos, J. Burdick, and A. Krause. Safe exploration for optimization with gaussian processes. In *International Conference on Machine Learning*, pages 997–1005. PMLR, 2015.

Y. Sui, J. Burdick, Y. Yue, et al. Stagewise safe bayesian optimization with gaussian processes. In *International Conference on Machine Learning*, pages 4781–4789. PMLR, 2018.

Y. Sui, V. Zhuang, J. W. Burdick, and Y. Yue. Stagewise safe bayesian optimization with gaussian processes, 2020.

J. Suk and S. Kpotufe. Tracking most severe arm changes in bandits. *arXiv preprint arXiv:2112.13838*, 2021.

S. Takemori. Distributionally-aware kernelized bandit problems for risk aversion. In *International Conference on Machine Learning*, pages 20933–20959. PMLR, 2022.

K. Takemura, S. Ito, D. Hatano, H. Sumita, T. Fukunaga, N. Kakimura, and K. ichi Kawarabayashi. A parameter-free algorithm for misspecified linear contextual bandits. In *International Conference on Artificial Intelligence and Statistics*, 2021.

D. Tang, R. Jain, A. Nayyar, and P. Nuzzo. Pure exploration for constrained best mixed arm identification with a fixed budget. *arXiv preprint arXiv:2405.15090*, 2024.

C. Tao, S. Blanco, and Y. Zhou. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pages 4877–4886, 2018.

P. S. Thomas, B. Castro da Silva, A. G. Barto, S. Giguere, Y. Brun, and E. Brunskill. Preventing undesirable behavior of intelligent machines. *Science*, 366(6468):999–1004, 2019.

A. Tirinzoni, M. Pirotta, M. Restelli, and A. Lazaric. An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. *arXiv preprint arXiv:2010.12247*, 2020.

M. J. Todd. *Minimum-volume ellipsoids: Theory and algorithms*. SIAM, 2016.

S. Vakili, N. Bouziani, S. Jalali, A. Bernacchia, and D.-s. Shiu. Optimal order simple regret for gaussian process bandits. *Advances in Neural Information Processing Systems*, 34, 2021a.

S. Vakili, J. Scarlett, and T. Javidi. Open problem: Tight online confidence intervals for rkhs elements. In *Conference on Learning Theory*, pages 4647–4652. PMLR, 2021b.

M. Valko, N. Korda, R. Munos, I. Flaounas, and N. Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

R. Vershynin. Introduction to the non-asymptotic analysis of random matrices, 2011.

A. Wagenmaker and D. J. Foster. Instance-optimality in interactive decision making: Toward a non-asymptotic theory. *arXiv preprint arXiv:2304.12466*, 2023.

A. Wagenmaker, J. Katz-Samuels, and K. Jamieson. Experimental design for regret minimization in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3088–3096. PMLR, 2021.

A. Wagenmaker, Y. Chen, M. Simchowitz, S. S. Du, and K. Jamieson. Reward-free rl is no harder than reward-aware rl in linear markov decision processes. *arXiv preprint arXiv:2201.11206*, 2022a.

A. J. Wagenmaker, Y. Chen, M. Simchowitz, S. Du, and K. Jamieson. First-order regret in reinforcement learning with linear function approximation: A robust estimation approach. In *International Conference on Machine Learning*, pages 22384–22429. PMLR, 2022b.

Y. Wang and A. Singh. Noise-adaptive margin-based active learning and lower bounds under tsybakov noise condition, 2015.

Z. Wang, A. J. Wagenmaker, and K. Jamieson. Best arm identification with safety constraints. In *International Conference on Artificial Intelligence and Statistics*, pages 9114–9146. PMLR, 2022.

C.-Y. Wei and H. Luo. Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach. In *Conference on Learning Theory*, pages 4300–4354. PMLR, 2021.

C.-Y. Wei, H. Luo, and A. Agarwal. Taking a hint: How to leverage loss predictors in contextual bandits?, 2020.

B. Woodworth, S. Gunasekar, M. I. Ohannessian, and N. Srebro. Learning non-discriminatory predictors, 2017.

J. Wu and P. I. Frazier. The parallel knowledge gradient method for batch bayesian optimization, 2018.

Y. Wu, Z. Zheng, G. Zhang, Z. Zhang, and C. Wang. Non-stationary a/b tests. 2022.

M. Xiao and Y. Guo. Online active learning for cost sensitive domain adaptation, 2013.

Z. Xiong, R. Camilleri, M. Fazel, L. Jain, and K. Jamieson. A/b testing and best-arm identification for linear bandits with robustness to non-stationarity. In *International Conference on Artificial Intelligence and Statistics*, pages 1585–1593. PMLR, 2024.

L. Xu, J. Honda, and M. Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851. PMLR, 2018.

Yahoo! Yahoo! webscope dataset ydata-frontpage-todaymodule-clicks-v1_0, 2011. URL `https://webscope.sandbox.yahoo.com/catalog.php?datatype=r`.

J. Yang and V. Tan. Minimax optimal fixed-budget best arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 35:12253–12266, 2022.

J. Yang and V. Y. Tan. Towards minimax optimal best arm identification in linear bandits. *arXiv e-prints*, pages arXiv–2105, 2021.

M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi. Fairness beyond disparate treatment and disparate impact. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, apr 2017a. doi: 10.1145/3038912.3052660. URL `https://doi.org/10.1145%2F3038912.3052660`.

M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi. Fairness constraints: Mechanisms for fair classification, 2017b.

A. Zanette, J. Zhang, and M. J. Kochenderfer. Robust super-level set estimation using gaussian processes. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 276–291. Springer, 2018.

H. Zenati, A. Bietti, E. Diemert, J. Mairal, M. Martin, and P. Gaillard. Efficient kernelized ucb for contextual bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 5689–5720. PMLR, 2022.

B. H. Zhang, B. Lemoine, and M. Mitchell. Mitigating unwanted biases with adversarial learning. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 335–340, 2018.

X. Zhou and B. Ji. On kernelized multi-armed bandits with constraints. *Advances in neural information processing systems*, 35:14–26, 2022.

Y. Zhu, D. Zhou, R. Jiang, Q. Gu, R. Willett, and R. Nowak. Pure exploration in kernel and neural bandits. *Advances in neural information processing systems*, 34:11618–11630, 2021.

Appendix A

Appendix for Chap. 2

Outline

The appendix is organized as follows. We first provide the proofs for the concentration bound of RIPS (Theorem 1), the computation of the inverse of the bilinear form (Lemma 1), the guarantees of the PTR procedure (Theorem 2), the regret bound of the RIPS regret minimization algorithm (Theorem 3), the sample complexity of the RIPS pure exploration algorithm (Theorem 4). We also establish the regret bound and the sample complexity guarantees of the PTR procedure. Then, we provide the proofs of the comparison of our variance term $f(\mathcal{X}, 1/T)$ with the information gain of (Srinivas et al., 2009) (Lemma 2) and with the effective dimension of (Alaoui and Mahoney, 2014) (Lemma 3) and prove a corollary of Theorem 1 of (Degenne et al., 2020). Last, we complete the details of the experiments.

A.1 Concentration of RIPS, Proof of Theorem 1

Proof. First note that

$$\begin{aligned}
\max_{v\in\mathcal{V}} \frac{|\langle\widehat{\theta}, v\rangle - \langle\theta_*, v\rangle|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} &= \max_{v\in\mathcal{V}} \frac{|\langle\widehat{\theta}, v\rangle - W^{(v)} + W^{(v)} - \langle\theta_*, v\rangle|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} \\
&\leq \max_{v\in\mathcal{V}} \frac{|\langle\widehat{\theta}, v\rangle - W^{(v)}|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} + \max_{v\in\mathcal{V}} \frac{|W^{(v)} - \langle\theta_*, v\rangle|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} \\
&= \min_{\theta} \max_{v\in\mathcal{V}} \frac{|\langle\theta, \phi(v)\rangle - W^{(v)}|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} + \max_{v\in\mathcal{V}} \frac{|W^{(v)} - \langle\theta_*, v\rangle|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}} \\
&\leq 2 \max_{v\in\mathcal{V}} \frac{|\langle\theta_*, v\rangle - W^{(v)}|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}}
\end{aligned}$$

which completes the second part of the lemma, so it suffices to show that each $|\langle\theta_*, v\rangle - W^{(v)}|$ is small.

We begin by bounding the variance of $v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t$ for any $t \in \mathbb{N}$ which is necessary to use the robust estimator. Note that

$$\mathbb{V}\text{ar}(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t) = \mathbb{E}[(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t)^2] - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t]^2$$

which means we can drop the second term to bound the variance by

$$\mathbb{E}[\Big(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\Big)^2] = \mathbb{E}[\Big(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)(\phi(x_t)^\top\theta_* + \eta_t + \xi_t)\Big)^2]$$

$$
\begin{aligned}
&= \mathbb{E}[\Big(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)(\phi(x_t)^\top\theta_* + \eta_t)\Big)^2] + \mathbb{E}[\Big(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)\Big)^2\xi_t^2] \\
&\le B^2\mathbb{E}[\Big(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)\Big)^2] + \sigma^2\mathbb{E}[\Big(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)\Big)^2] \\
&= (B^2+\sigma^2)\mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)\phi(x_t)^\top A^{(\gamma)}(\lambda)^{-1}v] \\
&\le (B^2+\sigma^2)\|v\|^2_{A^{(\gamma)}(\lambda)^{-1}}.
\end{aligned}
$$

Recalling that

$$
|\widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^T) - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1]| \le c_0\sqrt{\mathbb{V}\text{ar}(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1)\frac{\log(\frac{2}{\delta})}{T}}
$$

we have

$$
\begin{aligned}
|\langle\theta_*, v\rangle - W^{(v)}| &= |\langle\theta_*, v\rangle - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1] + \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1] - W^{(v)}| \\
&\le |\langle\theta_*, v\rangle - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1]| + |\widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^T) - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1]|.
\end{aligned}
$$

We now recall that $y_t = \langle\phi(x_t), \theta_*\rangle + \xi_t + \eta_{x_t}$ where ξ_t is a mean-zero, independent random variable with variance σ^2, and $|\eta_{x_t}| \le h$. Thus,

$$
\begin{aligned}
|\langle\theta_*, v\rangle - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1]| &= |\langle\theta_*, v\rangle - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\phi(x_1)^\top\theta_*] - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\eta_{x_1}]| \\
&\le |\langle\theta_*, v\rangle - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\phi(x_1)^\top\theta_*]| + |\mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\eta_{x_1}]|
\end{aligned}
$$

which we bound separately. Firstly,

$$
\begin{aligned}
|\langle\theta_*, v\rangle - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\phi(x_1)^\top\theta_*]| &= |\langle\theta_*, v\rangle - v^\top A^{(\gamma)}(\lambda)^{-1}A(\lambda)\theta_*| \\
&= \gamma|v^\top A^{(\gamma)}(\lambda)^{-1}\theta_*| \\
&= \gamma^{1/2}|v^\top(A(\lambda)+\gamma I)^{-1/2}(A(\lambda)/\gamma + I)^{-1/2}\theta_*| \\
&\le \gamma^{1/2}|v^\top(A(\lambda)+\gamma I)^{-1/2}\theta_*| \\
&\le \gamma^{1/2}\|v\|_{A^{(\gamma)}(\lambda)^{-1}}\|\theta_*\|
\end{aligned}
$$

and secondly,

$$
\begin{aligned}
|\mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\eta_{x_1}]| &\le \mathbb{E}[|v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\eta_{x_1}|] \\
&\le h\sqrt{\mathbb{E}[|v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)|^2]} \\
&= h\sqrt{v^\top A^{(\gamma)}(\lambda)^{-1}A(\lambda)A^{(\gamma)}(\lambda)^{-1}v} \\
&\le h\|v\|_{A^{(\gamma)}(\lambda)^{-1}}.
\end{aligned}
$$

Thus, putting it all together we have

$$
|\langle\theta_*, v\rangle - W^{(v)}| \le (\sqrt{\gamma}\|\theta_*\|_2 + h + c_0\sqrt{(B^2+\sigma^2)\frac{\log(2/\delta)}{T}})\|v\|_{A^{(\gamma)}(\lambda)^{-1}}.
$$

Union bounding over all $v \in \mathcal{V}$ completes the proof. □

A.2 Inverses and bilinear forms, Proof of Lemma 1

Proof of Proposition 1. The following manipulations are well-known, but we include them from completeness. Define

$$\Phi := [\phi(x_1)^\top, \dots, \phi(x_\tau)^\top]^\top$$

Holds

$$\Phi^\top(\Phi\Phi^\top + \gamma I) = (\Phi^\top\Phi + \gamma I)\Phi^\top \ .$$

And thus

$$(\Phi^\top\Phi + \gamma I)^{-1}\Phi^\top = \Phi^\top(\Phi\Phi^\top + \gamma I)^{-1} \ .$$

Now we use the expansion

$$\left(\Phi^\top\Phi + \gamma I\right) a = \Phi^\top\Phi a + \gamma a$$

to write

$$\begin{aligned} a &= \left(\Phi^\top\Phi + \gamma I\right)^{-1}\left(\Phi^\top\Phi a + \gamma a\right) \\ &= \left(\Phi^\top\Phi + \gamma I\right)^{-1}\Phi^\top\Phi a + \left(\Phi^\top\Phi + \gamma I\right)^{-1}\gamma a \\ &= \Phi^\top\left(\Phi\Phi^\top + \gamma I\right)^{-1}\Phi a + \gamma\left(\Phi^\top\Phi + \gamma I\right)^{-1} a \ . \end{aligned}$$

Then multiplying on the left side by $b^\top$ leads to

$$b^\top a = b^\top\Phi^\top\left(\Phi\Phi^\top + \gamma I\right)^{-1}\Phi a + \gamma b^\top\left(\Phi^\top\Phi + \gamma I\right)^{-1} a \ .$$

So

$$\begin{aligned} b^\top\left(\sum_{x'\in\mathcal{X}}\phi(x')\phi(x')^\top + \gamma I\right)^{-1} a &= \frac{1}{\gamma}b^\top a - \frac{1}{\gamma}b^\top\Phi^\top\left(\Phi\Phi^\top + \gamma I\right)^{-1}\Phi a \\ &= \frac{1}{\gamma}a^\top b - \frac{1}{\gamma}k(a)^\top\left(K + \gamma I\right)^{-1}k(b) \ . \end{aligned}$$

We now simply repeat with the same calculations with

$$\Phi_\lambda := [\sqrt{\lambda_1}\phi(x_1)^\top, \dots, \sqrt{\lambda_\tau}\phi(x_\tau)^\top]^\top \ ,$$

$$K_\lambda = \Phi_\lambda\Phi_\lambda^\top = \left[\sqrt{\lambda_i}\sqrt{\lambda_j}\phi(x_i)^\top\phi(x_j)\right]_{1\le i,j\le\tau} \ ,$$

and

$$k_\lambda(x) := \Phi_\lambda\phi(x) \in \mathbb{R}^\tau \ .$$

□

A.3 Guarantees of the PTR procedure, Proof of Theorem 2

We establish the proof in a finite dimension case where ϕ is the identity map and then extend it to any feature map ϕ. In both cases, we fix $\mathcal{X} \subset \mathbb{R}^d$ and consider $\lambda \in \triangle_{\mathcal{X}}$ to be the design we wish to round.

A.3.1 Finite dimension

Lemma 5. *Let $VDV^\top$ be the eigenvalue decomposition of the matrix $\sum_{x\in\mathcal{X}} \lambda_x xx^\top$, and denote $D = diag(d_1, \dots, d_d)$. For any $z \in \mathcal{V}$, as long as $\tau = \Omega(k/\epsilon)$, we can find an allocation $\{\widetilde{x}_i\}_{i=1}^{\tau} \subset \mathcal{X}$ such that*

$$z^\top \left(\sum_{i=1}^{\tau} \widetilde{x}_i \widetilde{x}_i^\top + \tau\gamma I_d \right)^{-1} z \le \max\{1+\epsilon, 2\} z^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x xx^\top + \tau\gamma I_d \right)^{-1} z \,,$$

where we defined $k = \max\{i : d_i \ge \gamma\}$.

Proof. Start by also denoting $V = [v_1, \dots, v_d]$. Then

$$\begin{aligned} z^\top (\tau \sum_{x\in\mathcal{X}} \lambda_x xx^\top + \tau\gamma I)^{-1} z &= z^\top (\tau V D V^\top + \tau\gamma I)^{-1} z \\ &= z^\top (\tau V D V^\top + \tau\gamma V V^\top)^{-1} z \\ &= z^\top V (\tau D + \tau\gamma I)^{-1} V^\top z \\ &= z^\top \left(\sum_{i=1}^{d} v_i v_i^\top \frac{1}{\tau d_i + \tau\gamma} \right) z \end{aligned}$$

Now, for any $k = \max\{i : d_i \ge \gamma\}$ we have

$$\begin{aligned} z^\top \left(\sum_{i=1}^{d} v_i v_i^\top \frac{1}{\tau d_i + \tau\gamma} \right) z &= z^\top \left(\sum_{i=1}^{k} v_i v_i^\top \frac{1}{\tau d_i + \tau\gamma} \right) z + z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau d_i + \tau\gamma} \right) z \\ &\ge z^\top \left(\sum_{i=1}^{k} v_i v_i^\top \frac{1}{\tau d_i + \tau\gamma} \right) z + \frac{1}{2} z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma} \right) z \\ &= (V_k^\top z)^\top V_k^\top \left(\sum_{i=1}^{k} v_i v_i^\top \frac{1}{\tau d_i + \tau\gamma} \right) V_k (V_k^\top z) + \frac{1}{2} z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma} \right) z \\ &= (V_k^\top z)^\top V_k^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x xx^\top + \tau\gamma I_d \right)^{-1} V_k (V_k^\top z) + \frac{1}{2} z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma} \right) z \\ &= (V_k^\top z)^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x (V_k^\top x)(V_k^\top x)^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z) + \frac{1}{2} z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma} \right) z. \end{aligned}$$

where we denote V_k and V_{-k} as the top k and bottom $d-k$ eigenvectors, respectively. But now we notice that for this first term, we have $\max\{\dim(\text{span}(\{V_k^\top z\}_{z\in\mathcal{V}})), \dim(\text{span}(\{V_k^\top x\}_{x\in\mathcal{X}}))\} \le k$ which now means that thanks to (Allen-Zhu et al., 2017) we can find an allocation $\{\widetilde{x}_i\}_{i=1}^{\tau} \subset \mathcal{X}$ such that

$$(V_k^\top z)^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x (V_k^\top x)(V_k^\top x)^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z) \ge \frac{1}{1+\epsilon} (V_k^\top z)^\top \left(\sum_{i=1}^{\tau} (V_k^\top \widetilde{x})(V_k^\top \widetilde{x})^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z)$$

as long as $\tau = \Omega(k/\epsilon)$. Putting it altogether we have

$$
\begin{aligned}
&z^\top \left(\sum_{i=1}^{\tau} \widetilde{x}_i \widetilde{x}_i^\top + \tau\gamma I_d \right)^{-1} z \\
&= (V_k^\top z)^\top \left(\sum_{i=1}^{\tau} (V_k^\top \widetilde{x}_i)(V_k^\top \widetilde{x}_i)^\top + \tau\gamma V_k^\top V_k \right)^{-1} (V_k^\top z) + (V_{-k}^\top z)^\top \left(\sum_{i=1}^{\tau} (V_{-k}^\top \widetilde{x}_i)(V_{-k}^\top \widetilde{x}_i)^\top + \tau\gamma V_{-k}^\top V_{-k} \right)^{-1} (V_{-k}^\top z) \\
&\leq (V_k^\top z)^\top \left(\sum_{i=1}^{\tau} (V_k^\top \widetilde{x}_i)(V_k^\top \widetilde{x}_i)^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z) + (V_{-k}^\top z)^\top \left(\tau\gamma V_{-k}^\top V_{-k} \right)^{-1} (V_{-k}^\top z) \\
&= (V_k^\top z)^\top \left(\sum_{i=1}^{\tau} (V_k^\top \widetilde{x}_i)(V_k^\top \widetilde{x}_i)^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z) + z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma} \right) z \\
&\leq (V_k^\top z)^\top \left(\sum_{i=1}^{\tau} (V_k^\top \widetilde{x}_i)(V_k^\top \widetilde{x}_i)^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z) + 2z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma + \tau d_i} \right) z \\
&\leq (1+\epsilon)(V_k^\top z)^\top \left(\tau \sum_{x \in \mathcal{X}} \lambda_x (V_k^\top x)(V_k^\top x)^\top + \tau\gamma I_k \right)^{-1} (V_k^\top z) + 2z^\top \left(\sum_{i=k+1}^{d} v_i v_i^\top \frac{1}{\tau\gamma + \tau d_i} \right) z \\
&\leq \max\{1+\epsilon, 2\} z^\top \left(\tau \sum_{x \in \mathcal{X}} \lambda_x x x^\top + \tau\gamma I_d \right)^{-1} z.
\end{aligned}
$$

□

Oftentimes k can be much smaller than $\min\{d, |\mathcal{X}|\}$, especially for large γ. For instance, for $\mathcal{X} = \mathcal{Z} = \{a\mathbf{e}_1\} \cup \{\mathbf{e}_i : i \in [d]\}$ with $a \gg \gamma = 1$, even as $d \to \infty$ we have that $k = 1$ since λ^* will be the majority of its mass on $\mathbf{e}_1$.

A.3.2 Connection to kernels

We now get back to our initial setting. Consider K the kernel matrix of $\mathcal{X} = \{x_1, \ldots, x_n\}$. Take $\widetilde{\Phi} \in \mathbb{R}^{n \times n}$ such that $K = \widetilde{\Phi}\widetilde{\Phi}^\top$ (can easily done by diagonalizing K). Consider the rows of $\widetilde{\Phi}$ and name these $\widetilde{\phi}(x_i)$. Then, we have by definition $\phi(x_i)^\top \phi(x_j) = [K]_{ij}$ and we have by construction $\widetilde{\phi}(x_i)^\top \widetilde{\phi}(x_j) = [K]_{ij}$, which importantly leads to $\phi(x_i)^\top \phi(x_j) = \widetilde{\phi}(x_i)^\top \widetilde{\phi}(x_j)$.
Fix $v \in \mathcal{V} \subset \mathcal{X}$. We have from Lemma 1

$$
\phi(v)^\top \left(\sum_{i=1}^{\tau} \phi(x_i)\phi(x_i)^\top + \tau\gamma I \right)^{-1} \phi(v) = \phi(v)^\top \phi(v)/(\tau\gamma) - \phi(v)^\top \Phi^\top (\Phi\Phi^\top + \tau\gamma I_n)^{-1} \Phi\phi(v)/(\tau\gamma) \, .
$$

This only involves scalar products of the form $\phi(x_i)^\top \phi(x_j)$, such that the property $\phi(x_i)^\top \phi(x_j) = \widetilde{\phi}(x_i)^\top \widetilde{\phi}(x_j)$ allows us to write the variance as

$$
\phi(v)^\top \left(\sum_{i=1}^{\tau} \phi(x_i)\phi(x_i)^\top + \tau\gamma I \right)^{-1} \phi(v) = \phi(v)^\top \phi(v)/(\tau\gamma) - \phi(v)^\top \Phi^\top (\Phi\Phi^\top + \tau\gamma I_n)^{-1} \Phi\phi(v)/(\tau\gamma)
$$

$$= \widetilde{\phi}(v)^\top \widetilde{\phi}(v)/(\tau\gamma) - \widetilde{\phi}^\top \widetilde{\Phi}^\top (\widetilde{\Phi}\widetilde{\Phi}^\top + \tau\gamma I_n)^{-1} \widetilde{\Phi}\widetilde{\phi}(v)/(\tau\gamma)$$
$$= \widetilde{\phi}(v)^\top \left(\sum_{i=1}^{\tau} \widetilde{\phi}(x_i)\widetilde{\phi}(x_i)^\top + \tau\gamma I_n \right)^{-1} \widetilde{\phi}(v) \ .$$

The same trick allows us to write

$$\phi(v)^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \tau\gamma I \right)^{-1} \phi(v) = \widetilde{\phi}(v)^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x \widetilde{\phi}(x)\widetilde{\phi}(x)^\top + \tau\gamma I_n \right)^{-1} \widetilde{\phi}(v) \ .$$

Let $V\Delta V^\top$ be the eigenvalue decomposition of the matrix $\sum_{x\in\mathcal{X}} \lambda_x \widetilde{\phi}(x)\widetilde{\phi}(x)^\top$, and denote $\Delta = \mathrm{diag}(d_1, \dots, d_n)$. We know from lemma 5 that with $\tau = \Omega(\widetilde{d}(\lambda,\gamma)/\epsilon)$ and $\widetilde{d}(\lambda,\gamma) = \max\{i : d_i \geq \gamma\}$ we can find an allocation $\{\widetilde{x}_i\}_{i=1}^{\tau} \subset \mathcal{X}$ such that

$$\widetilde{\phi}(v)^\top \left(\sum_{i=1}^{\tau} \widetilde{\phi}(\widetilde{x}_i)\widetilde{\phi}(\widetilde{x}_i) + \tau\gamma I_n \right)^{-1} \widetilde{\phi}(v) \leq \max\{1+\epsilon, 2\} \widetilde{\phi}(v)^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x \widetilde{\phi}(x)\widetilde{\phi}(x)^\top + \tau\gamma I_n \right)^{-1} \widetilde{\phi}(v) \ ,$$

which yields to the following result.
For any $v \in \mathcal{V} \subset \mathcal{X}$, as long as $\tau = \Omega(\widetilde{d}(\lambda,\gamma)/\epsilon)$, we can find an allocation $\{\widetilde{x}_i\}_{i=1}^{\tau} \subset \mathcal{X}$ such that

$$\phi(v)^\top \left(\sum_{i=1}^{\tau} \phi(\widetilde{x}_i)\phi(\widetilde{x}_i)^\top + \tau\gamma I_d \right)^{-1} \phi(v) \leq \max\{1+\epsilon, 2\} \phi(v)^\top \left(\tau \sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \tau\gamma I_d \right)^{-1} \phi(v)$$

and $\widetilde{d}(\lambda,\gamma) = \max\{i : d_i \geq \gamma\}$.

And we can take the suppremum over $v \in \mathcal{V}$ to get to the result of Theorem 2.

A.4 Main regret argument, Proof of Theorem 3

In this section, we can consider without loss of generality that ϕ is the identity map. Indeed, the features of the actions - thus denoted x here and $\phi(x)$ in the rest of the paper - appear in this proof only through scalar products.

Define $f(\mathcal{V};\gamma) = \inf_{\lambda\in\triangle_\mathcal{V}} \max_{v\in\mathcal{V}} \|v\|^2_{(\sum_{y\in\mathcal{V}} \lambda_y yy^\top + \gamma I)^{-1}}$ and $\bar{f}(\mathcal{X};\gamma) := \max_{\mathcal{V}\subseteq\mathcal{X}} f(\mathcal{V};\gamma)$.

Define the event

$$\mathcal{E} := \bigcap_{\ell=1}^{\infty} \bigcap_{x\in\mathcal{X}_\ell} \left\{ |\langle x, \widehat{\theta}_\ell - \theta_* \rangle| \leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \right\}$$

Lemma 6. *We have* $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$.

Proof. For any $\mathcal{V} \subseteq \mathcal{X}$ and $x \in \mathcal{V}$ define

$$\mathcal{E}_{x,\ell}(\mathcal{V}) = \{ |\langle x, \widehat{\theta}_\ell(\mathcal{V}) - \theta_* \rangle| \leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \}$$

where $\widehat{\theta}_\ell(\mathcal{V})$ is the estimator that would be constructed by the algorithm at stage ℓ with $\mathcal{X}_\ell = \mathcal{V}$. For fixed $\mathcal{V} \subset \mathcal{X}$ and $\ell \in \mathbb{N}$ we apply Theorem 1 with $\tau = \tau_\ell$ so that with probability at least $1 - \frac{\delta}{2\ell^2|\mathcal{X}|}$ we have that for any $x \in \mathcal{V}$

$$\begin{aligned}
|\langle x, \widehat{\theta}_\ell(\mathcal{V}) - \theta_* \rangle| &\le \|x\|_{A^{(\gamma)}(\lambda_\ell)^{-1}} \left(\sqrt{\gamma}\|\theta_*\|_2 + h + c_0 \sqrt{(B^2 + \sigma^2) \frac{\log(4\ell^2|\mathcal{X}|/\delta)}{\tau_\ell}} \right) \\
&\le \sqrt{f(\mathcal{V};\gamma)} \left(\sqrt{\gamma}\|\theta_*\|_2 + h + \epsilon_\ell / \sqrt{f(\mathcal{V};\gamma)} \right) \\
&\le \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h) \sqrt{\bar{f}(\mathcal{X};\gamma)}
\end{aligned}$$

Noting that $\mathcal{E} := \bigcap_{\ell=1}^\infty \bigcap_{x \in \mathcal{X}_\ell} \mathcal{E}_{x,\ell}(\mathcal{X}_\ell)$ we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\ell=1}^\infty \bigcup_{x \in \mathcal{X}_\ell} \{\mathcal{E}_{x,\ell}^c(\mathcal{X}_\ell)\} \right) &\le \sum_{\ell=1}^\infty \mathbb{P}\left(\bigcup_{x \in \mathcal{X}_\ell} \{\mathcal{E}_{x,\ell}^c(\mathcal{X}_\ell)\} \right) \\
&= \sum_{\ell=1}^\infty \sum_{\mathcal{V} \subseteq \mathcal{X}} \mathbb{P}\left(\bigcup_{x \in \mathcal{V}} \{\mathcal{E}_{x,\ell}^c(\mathcal{V})\}, \mathcal{X}_\ell = \mathcal{V} \right) \\
&= \sum_{\ell=1}^\infty \sum_{\mathcal{V} \subseteq \mathcal{X}} \mathbb{P}\left(\bigcup_{x \in \mathcal{V}} \{\mathcal{E}_{x,\ell}^c(\mathcal{V})\} \right) \mathbb{P}(\mathcal{X}_\ell = \mathcal{V}) \\
&\le \sum_{\ell=1}^\infty \sum_{\mathcal{V} \subseteq \mathcal{X}} \tfrac{\delta}{2\ell^2|\mathcal{X}|} |\mathcal{V}| \mathbb{P}(\mathcal{X}_\ell = \mathcal{V}) \\
&\le \sum_{\ell=1}^\infty \sum_{\mathcal{V} \subseteq \mathcal{X}} \tfrac{\delta}{2\ell^2} \mathbb{P}(\widehat{\mathcal{X}}_\ell = \mathcal{V}) \le \delta
\end{aligned}$$

□

Lemma 7. *For all $\ell \in \mathbb{N}$ we have* $\max_{x \in \mathcal{X}_\ell} \mu_* - \mu_x \le \max\{16\epsilon_\ell, 32(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\}$.

Proof. An arm $x \in \mathcal{X}_\ell$ is discarded (i.e., not in $\mathcal{X}_{\ell+1}$) if $\max_{x' \in \mathcal{X}_\ell} \langle x', \widehat{\theta} \rangle - \langle x, \widehat{\theta} \rangle > 4\epsilon_\ell$. Let $\bar{\ell} := \max\{\ell : \epsilon_\ell > 2(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\}$. If $x_* = \arg\max_{x \in \mathcal{X}} \mu_x$ then $x_* \in \mathcal{X}_1$. Now if $x_* \in \mathcal{X}_\ell$ for some $\ell \le \bar{\ell}$, then for any $x' \in \mathcal{X}_\ell$ we have

$$\begin{aligned}
\langle x', \widehat{\theta} \rangle - \langle x_*, \widehat{\theta} \rangle &\le \langle x' - x_*, \theta_* \rangle + 2\epsilon_\ell + 2(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\le \mu_x - \mu_{x_*} + 2h + 2\epsilon_\ell + 2(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\le 2\epsilon_\ell + 4(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\le 4\epsilon_\ell
\end{aligned}$$

which implies $x_* \in \mathcal{X}_{\ell+1}$. Moreover, suppose that $\ell \le \bar{\ell}$ and there exists some $x \in \mathcal{X}_\ell$ such that $\mu_* - \mu_x > 8\epsilon_\ell$, then

$$\max_{x' \in \mathcal{X}_\ell} \langle x', \widehat{\theta} \rangle - \langle x, \widehat{\theta} \rangle \ge \langle x_*, \widehat{\theta} \rangle - \langle x, \widehat{\theta} \rangle$$

$$
\begin{aligned}
&\geq \langle x_* - x, \theta_* \rangle - 2\epsilon_\ell - 2(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\geq \mu_* - \mu_x - 2h - 2\epsilon_\ell - 2(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\geq \mu_* - \mu_x - 2\epsilon_\ell - 4(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\geq \mu_* - \mu_x - 4\epsilon_\ell \\
&> 4\epsilon_\ell
\end{aligned}
$$

which implies $\max_{x \in \mathcal{X}_{\ell+1}} \mu_* - \mu_x \leq 8\epsilon_\ell = 16\epsilon_{\ell+1}$. Because $\mathcal{X}_{\ell+1} \subseteq \mathcal{X}_\ell$ we have for $\ell > \bar{\ell}$ that

$$
\begin{aligned}
\max_{x \in \mathcal{X}_\ell} \mu_* - \mu_x &\leq \max_{x \in \mathcal{X}_{\bar{\ell}+1}} \mu_* - \mu_x \\
&\leq 16\epsilon_{\bar{\ell}+1} \\
&\leq 32(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)}.
\end{aligned}
$$

Thus, $\max_{x \in \mathcal{X}_\ell} \mu_* - \mu_x \leq \max\{16\epsilon_\ell, 32(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\}$. □

We now compute the final regret bound. After T steps of the algorithm, let T_x denote the number of times arm x is played. Let $\Gamma = (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)}$. If L is the final round reached after T steps, we have

$$
\begin{aligned}
\sum_{x \in \mathcal{X}} (\mu_* - \mu_x) T_x &\leq \sum_{\ell=1}^{L} \max_{x \in \mathcal{X}_\ell} (\mu_* - \mu_x)\tau_\ell \\
&\leq \sum_{\ell=1}^{L} \tau_\ell \max\{16\epsilon_\ell, 32(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\} \\
&\leq \sum_{\ell=1}^{L} \tau_\ell \max\{16\epsilon_\ell, 32\Gamma\} \\
&\leq \sum_{\ell:\epsilon_\ell < 2\Gamma} 32\Gamma\tau_\ell + \sum_{\ell:\epsilon_\ell \geq 2\Gamma} \epsilon_\ell \tau_\ell \\
&\leq \sum_{\ell:\epsilon_\ell < 2\Gamma} 32\Gamma\tau_\ell + 16\nu T + \sum_{\ell:\epsilon_\ell \geq 2\Gamma \vee \nu} 16\epsilon_\ell \tau_\ell \\
&\leq \sum_{\ell:\epsilon_\ell < 2\Gamma} 32\Gamma\tau_\ell + 16\nu T + \sum_{\ell:\epsilon_\ell \geq \nu} 16\epsilon_\ell \tau_\ell \\
&\leq c\left(\Gamma T + \nu T + \sum_{\ell:\epsilon_\ell \geq \nu} \epsilon_\ell \cdot \left(c_0(B^2 + \sigma^2)\epsilon_\ell^{-2} f(\mathcal{X}_\ell;\gamma)\log(4\ell^2|\mathcal{X}|/\delta) + c_1 \log(|\mathcal{X}|/\delta)\right)\right) \\
&\leq c\left(\Gamma T + \nu T + \cdot \left(c_0(B^2 + \sigma^2) f(\mathcal{X}_\ell;\gamma)\log(4\ell^2|\mathcal{X}|/\delta) + c_1 \log(|\mathcal{X}|/\delta)\right) \sum_{\ell:\epsilon_\ell \geq \nu} \epsilon_\ell\right) \\
&\leq c\left(\Gamma T + \nu T + \nu^{-1} c_0(B^2 + \sigma^2)\bar{f}(\mathcal{X};\gamma)\log(4\lceil\log_2(1/\nu)\rceil^2|\mathcal{X}|/\delta) + c_1\log(|\mathcal{X}|/\delta)\right).
\end{aligned}
$$

Choosing $\nu = \sqrt{c_0(B^2 + \sigma^2)\bar{f}(\mathcal{X};\gamma)\log(|\mathcal{X}|/\delta)/T}$ and plugging Γ back in yields

$$
\sum_{x \in \mathcal{X}} (\mu_* - \mu_x) T_x \leq c'\sqrt{\bar{f}(\mathcal{X};\gamma)}\left(T(\sqrt{\gamma}\|\theta_*\|_2 + h) + \sqrt{(B^2 + \sigma^2)\log(|\mathcal{X}|\log(T)/\delta)T}\right) + c_1 \log(|\mathcal{X}|/\delta).
$$

Choosing $\gamma = 1/T$ yields

$$\sum_{x \in \mathcal{X}} (\mu_* - \mu_x) T_x \leq c' \sqrt{\bar{f}(\mathcal{X}; 1/T)} \left(hT + \sqrt{(\|\theta_*\|^2 + \max_{x \in \mathcal{X}} \langle x, \theta_* \rangle + \sigma^2) \log(|\mathcal{X}| \log(T)/\delta) T} \right) + c_1 \log(|\mathcal{X}|/\delta).$$

A.5 Main robust pure exploration result, Proof of Theorem 4

For any $\mathcal{V} \subset \mathcal{Z}$ define $f(\mathcal{X}, \mathcal{V}; \gamma) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{v, v' \in \mathcal{V}} \|v - v'\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top + \gamma I)^{-1}}$

Lemma 8. *Define*

$$\bar{\epsilon} = 8 \min\{\epsilon \geq 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \leq \epsilon\}; \gamma)}) \leq \epsilon\}.$$

Then $\max_{z \in \mathcal{Z}_\ell} \mu_* - \mu_z \leq 8 \max\{\epsilon_\ell, \bar{\epsilon}\}$ *for all* $\ell \geq 0$ *with probability at least* $1 - \delta$.

We use Theorem 1 again, with here $\mathcal{V} \subset \mathcal{Z} \subset \mathbb{R}^d$:

$$\max_{v \in \mathcal{V}} \frac{|\langle \theta_*, \phi(v) \rangle - W^{(\phi(v))}|}{\|\phi(v)\|_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top + \gamma I)^{-1}}} \leq \sqrt{\gamma}\|\theta_*\|_2 + h + c_0 \sqrt{\frac{(B^2 + \sigma^2)}{T} \log(2|\mathcal{V}|/\delta)}$$

which motivates the choice

$$\tau_\ell = c_0^2 \epsilon_\ell^{-2} f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)(B^2 + \sigma^2) \log(2\ell^2 |\mathcal{Z}|^2/\delta)$$

Define the event

$$\mathcal{E} := \bigcap_{\ell=1}^{\infty} \bigcap_{z, z' \in \mathcal{Z}_\ell} \left\{ |\langle \phi(z) - \phi(z'), \widehat{\theta}_\ell - \theta_* \rangle| \leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h) \sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)} \right\}$$

Lemma 9. *We have* $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$.

Proof. This proof follows the analogous result for regret almost identically. We include it for completeness. For any $\mathcal{V} \subseteq \mathcal{Z}$ and $x \in \mathcal{X}$ define

$$\mathcal{E}_{z, z', \ell}(\mathcal{V}) = \{|\langle \phi(z) - \phi(z'), \widehat{\theta}_\ell(\mathcal{V}) - \theta_* \rangle| \leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h) \sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)}\}$$

where $\widehat{\theta}_\ell(\mathcal{V})$ is the estimator that would be constructed by the algorithm at stage ℓ with $\mathcal{Z}_\ell = \mathcal{V}$. For fixed $\mathcal{V} \subset \mathcal{Z}$ and $\ell \in \mathbb{N}$ we apply Theorem 1 with $T = \tau_\ell$ so that with probability at least $1 - \frac{\delta}{\ell^2 |\mathcal{Z}|^2}$ we have that for any $z, z' \in \mathcal{V}$

$$\begin{aligned}
|\langle \phi(z) - \phi(z'), \widehat{\theta}_\ell(\mathcal{V}) - \theta_* \rangle| &\leq \|\phi(z) - \phi(z')\|_{A^{(\gamma)}(\lambda_\ell)^{-1}} (\sqrt{\gamma}\|\theta_*\|_2 + h + c_0 \sqrt{(B^2 + \sigma^2) \frac{\log(2\ell^2 |\mathcal{Z}|^2/\delta)}{\tau_\ell}}) \\
&\leq \sqrt{f(\mathcal{X}, \mathcal{V}; \gamma)} \left(\sqrt{\gamma}\|\theta_*\|_2 + h + \epsilon_\ell / \sqrt{f(\mathcal{X}, \mathcal{V}; \gamma)} \right) \\
&\leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h) \sqrt{f(\mathcal{X}, \mathcal{V}; \gamma)}
\end{aligned}$$

Noting that $\mathcal{E} := \bigcap_{\ell=1}^{\infty} \bigcap_{z,z' \in \mathcal{Z}_\ell} \mathcal{E}_{z,z',\ell}(\mathcal{Z}_\ell)$ we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\ell=1}^{\infty} \bigcup_{z,z' \in \mathcal{Z}_\ell} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{Z}_\ell)\}\right) &\leq \sum_{\ell=1}^{\infty} \mathbb{P}\left(\bigcup_{z,z' \in \mathcal{Z}_\ell} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{Z}_\ell)\}\right) \\
&= \sum_{\ell=1}^{\infty} \sum_{\mathcal{V} \subseteq \mathcal{Z}} \mathbb{P}\left(\bigcup_{z,z' \in \mathcal{V}} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{V})\}, \mathcal{Z}_\ell = \mathcal{V}\right) \\
&= \sum_{\ell=1}^{\infty} \sum_{\mathcal{V} \subseteq \mathcal{Z}} \mathbb{P}\left(\bigcup_{z,z' \in \mathcal{V}} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{V})\}\right) \mathbb{P}(\mathcal{Z}_\ell = \mathcal{V}) \\
&\leq \sum_{\ell=1}^{\infty} \sum_{\mathcal{V} \subseteq \mathcal{Z}} \tfrac{\delta}{\ell^2 |\mathcal{Z}|^2} \binom{|\mathcal{V}|}{2} \mathbb{P}(\mathcal{Z}_\ell = \mathcal{V}) \\
&\leq \sum_{\ell=1}^{\infty} \sum_{\mathcal{V} \subseteq \mathcal{Z}} \tfrac{\delta}{2\ell^2} \mathbb{P}(\mathcal{Z}_\ell = \mathcal{V}) \leq \delta
\end{aligned}$$

□

Lemma 10. *Define* $S_1 = \mathcal{Z}$ *and* $S_{\ell+1} = \{z \in S_\ell : \langle \phi(z_*) - \phi(z), \theta_* \rangle \leq 3\epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, S_\ell; \gamma)}\}$. *Define*

$$\bar{\ell} = \max\{\ell : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, S_\ell; \gamma)}) \leq \epsilon_\ell\}.$$

For all $\ell \in \mathbb{N}$ *we have* $\max_{z \in \mathcal{Z}_\ell} \mu_* - \mu_z \leq 8 \max\{\epsilon_\ell, \epsilon_{\bar{\ell}}\}$.

Proof. An arm $z \in \mathcal{Z}_\ell$ is discarded (i.e., not in $\mathcal{Z}_{\ell+1}$) if $\max_{z' \in \mathcal{Z}_\ell} \langle \phi(z') - \phi(z), \widehat{\theta}_\ell \rangle > 2\epsilon_\ell$.

We will show $\{z_* \in \mathcal{Z}_\ell\} \cap \{\mathcal{Z}_\ell \subset S_\ell\} \cap \{\ell \leq \bar{\ell}\} \implies \{z_* \in \mathcal{Z}_{\ell+1}\} \cap \{\mathcal{Z}_{\ell+1} \subset S_{\ell+1}\}$. Noting that $\{z_* \in \mathcal{Z}_\ell\} \cap \{\mathcal{Z}_\ell \subset S_\ell\}$ holds for $\ell = 1$, we will assume an inductive hypothesis of this condition for some $\ell \leq \bar{\ell}$.

First we will show $\{z_* \in \mathcal{Z}_\ell\} \cap \{\mathcal{Z}_\ell \subset S_\ell\} \cap \{\ell \leq \bar{\ell}\} \implies \{z_* \in \mathcal{Z}_{\ell+1}\}$. On $\{z_* \in \mathcal{Z}_\ell\} \cap \{\mathcal{Z}_\ell \subset S_\ell\} \cap \{\ell \leq \bar{\ell}\}$, we have for any $z' \in \mathcal{Z}_\ell$ that

$$\begin{aligned}
\langle \phi(z') - \phi(z_*), \widehat{\theta}_\ell \rangle &\leq \langle \phi(z') - \phi(z_*), \theta_* \rangle + \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)} \\
&\leq \mu_{z'} - \mu_{z_*} + 2h + \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)} \\
&\leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)}) \\
&\leq \epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, S_\ell; \gamma)}) \\
&\leq 2\epsilon_\ell
\end{aligned}$$

which implies z_* is not eliminated, that is, $z_* \in \mathcal{Z}_{\ell+1}$. The second-to-last inequality follows from

$$\begin{aligned}
f(\mathcal{X}, \mathcal{Z}_\ell; \gamma) &= \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z,z' \in \mathcal{Z}_\ell} \|\phi(z) - \phi(z')\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}} \\
&\leq \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z,z' \in S_\ell} \|\phi(z) - \phi(z')\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}} \\
&= f(\mathcal{X}, S_\ell; \gamma).
\end{aligned}$$

Now we will show $\{z_* \in \mathcal{Z}_\ell\} \cap \{\mathcal{Z}_\ell \subset S_\ell\} \cap \{\ell \le \bar{\ell}\} \implies \{\mathcal{Z}_{\ell+1} \subset S_{\ell+1}\}$. For any $z \in \mathcal{Z}_\ell \cap S_{\ell+1}^c$ we have

$$\begin{aligned}
\max_{z' \in \mathcal{Z}_\ell} \langle \phi(z') - \phi(z), \widehat{\theta}_\ell \rangle &\ge \langle \phi(z_*) - \phi(z), \widehat{\theta}_\ell \rangle \\
&\ge \langle \phi(z_*) - \phi(z), \theta_* \rangle - \epsilon_\ell - (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)} \\
&> 3\epsilon_\ell + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, S_\ell; \gamma)} - \epsilon_\ell - (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)} \\
&\ge 2\epsilon_\ell
\end{aligned}$$

which implies $z \notin \mathcal{Z}_{\ell+1}$, and $\mathcal{Z}_{\ell+1} \subset S_{\ell+1}$.

Thus, for $\ell \le \bar{\ell}$ we have

$$\begin{aligned}
\max_{z \in \mathcal{Z}_\ell} \mu_* - \mu_z &\le \max_{z \in \mathcal{Z}_\ell} \langle \phi(z_*) - \phi(z), \theta_* \rangle + 2h \\
&\le \max_{z \in \mathcal{S}_\ell} \langle \phi(z_*) - \phi(z), \theta_* \rangle + 2h \\
&\le 3\epsilon_{\ell-1} + 2h + (\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, S_{\ell-1}; \gamma)} \\
&\le 3\epsilon_{\ell-1} + (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, S_{\ell-1}; \gamma)}) \\
&\le 4\epsilon_{\ell-1} = 8\epsilon_\ell.
\end{aligned}$$

And because $\mathcal{Z}_{\ell+1} \subseteq \mathcal{Z}_\ell$ we always have that $\max_{z \in \mathcal{Z}_\ell} \mu_* - \mu_z \le 8 \max\{\epsilon_\ell, \epsilon_{\bar{\ell}}\}$. Note that

$$\begin{aligned}
\bar{\ell} &= \max\{\ell : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, S_\ell; \gamma)}) \le \epsilon_\ell\} \\
&\ge \max\{\ell : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \le 4\epsilon_\ell\}; \gamma)}) \le \epsilon_\ell\} \\
&= \max\{\ell : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \le 4\epsilon_\ell\}; \gamma)}) \le 4\epsilon_\ell\} \\
&= -2 + \max\{\ell : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \le \epsilon_\ell\}; \gamma)}) \le \epsilon_\ell\} \\
&\ge -3 - \log_2(\min\{\epsilon > 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \le \epsilon\}; \gamma)}) \le \epsilon\})
\end{aligned}$$

which defines $\bar{\epsilon}$. □

The sample complexity to return an $8(\Delta \vee \bar{\epsilon})$-good arm is equal to

$$\begin{aligned}
&\sum_{\ell=1}^{\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil} \tau_\ell = \sum_{\ell=1}^{\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil} \left\lceil \max\left\{ c_1 \log(|\mathcal{Z}|/\delta), c_0 \epsilon_\ell^{-2} f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)(B^2 + \sigma^2) \log(2\ell^2 |\mathcal{Z}|^2/\delta) \right\} \right\rceil \\
&\le c \left(c_1 \log(|\mathcal{Z}|/\delta) \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) + c_0 (B^2 + \sigma^2) \log(2\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil^2 |\mathcal{Z}|^2/\delta) \sum_{\ell=1}^{\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil} \epsilon_\ell^{-2} f(\mathcal{X}, \mathcal{Z}_\ell; \gamma) \right) \\
&\le c \left(c_1 \log(|\mathcal{Z}|/\delta) \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) + c_0 (B^2 + \sigma^2) \log(2\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil^2 |\mathcal{Z}|^2/\delta) \sum_{\ell=1}^{\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil} \epsilon_\ell^{-2} f(\mathcal{X}, S_\ell; \gamma) \right) \\
&\le c \left(c_1 \log(|\mathcal{Z}|/\delta) \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) + 16 \lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil c_0 (B^2 + \sigma^2) \log(2\lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1}) \rceil^2 |\mathcal{Z}|^2/\delta) \rho^*(\gamma, \bar{\epsilon}) \right)
\end{aligned}$$

where the last line follows from

$$\rho^*(\gamma, \bar{\epsilon}) = \inf_{\lambda \in \Delta_\mathcal{X}} \sup_{z \in \mathcal{Z}} \frac{\|\phi(z_*) - \phi(z)\|^2_{\left(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I\right)^{-1}}}{\max\{\bar{\epsilon}^2, \langle \phi(z_*) - \phi(z), \theta_* \rangle^2)\}}$$

$$
\begin{aligned}
&= \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{\ell\le\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sup_{z\in S_\ell} \frac{\|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}}{\max\{\bar\epsilon^2,((\phi(z_*)-\phi(z))^\top\theta_*)^2\}} \\
&\ge \inf_{\lambda\in\Delta_{\mathcal{X}}} \frac{1}{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sup_{z\in S_\ell} \frac{\|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}}{\max\{\bar\epsilon^2,((\phi(z_*)-\phi(z))^\top\theta_*)^2\}} \\
&\ge \frac{1}{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{z\in S_\ell} \frac{\|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}}{\max\{\bar\epsilon^2,((\phi(z_*)-\phi(z))^\top\theta_*)^2\}} \\
&\ge \frac{1}{4\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} 2^{2\ell} \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{z\in S_\ell} \|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}} \\
&\ge \frac{1}{16\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} 2^{2\ell} \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{z,z'\in S_\ell} \|\phi(z)-\phi(z')\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}} \\
&= \frac{1}{16\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} 2^{2\ell} f(\mathcal{X},S_\ell;\gamma)\ .
\end{aligned}
$$

A.6 Proofs for the regret bound and the sample complexity of the alternative baseline

Note importantly that in this section the stochastic noise is sub-gaussian.

A.6.1 Concentration of the sparse estimator

Lemma 11. *Fix a finite set $\mathcal{V}\subset\mathcal{H}$, finite set $\mathcal{X}$, and let $\phi:\mathcal{X}\to\mathcal{H}$. Fix any $x_1,\dots,x_T$ and assume $y_t=\langle\phi(x_t),\theta_*\rangle+\xi_t+\eta_t$ where each ξ_t is independent, mean-zero, sub-gaussian with parameter σ^2, and each η_t satisfies $|\eta_t|\le h$. If $\widehat\theta$ is the regularized least squares estimator with regularization $\gamma>0$ then*

$$
\max_{v\in\mathcal{V}} \frac{|\langle\widehat\theta,v\rangle-\langle\theta_*,v\rangle|}{\|v\|_{(\sum_{i=1}^T\phi(x_i)\phi(x_i)^\top+\gamma I)^{-1}}} \le (\sqrt{\gamma}\|\theta_*\|_2+h\sqrt{T}+\sqrt{2\sigma^2\log(2|\mathcal{V}|/\delta)})
$$

with probability at least $1-\delta$.

Proof. Recall first the definition of regularized least squares estimator $\widehat\theta$:

$$
\widehat\theta = G_\gamma^{-1}\Phi^\top Y
$$

where $G_\gamma:=\sum_{i=1}^T\phi(x_i)\phi(x_i)^\top+\gamma I$. Defining $\xi:=(\xi_1,\dots,\xi_T)$ and $\eta:=(\eta_1,\dots,\eta_T)$, holds

$$
\begin{aligned}
v^\top(\widehat\theta-\theta_*) &= v^\top G_\gamma^{-1}\Phi^\top Y - v^\top\theta_* \\
&= v^\top G_\gamma^{-1}\Phi^\top(\Phi\theta_*+\xi+\eta)-v^\top\theta_* \\
&= v^\top G_\gamma^{-1}\Phi^\top\Phi\theta_* - v^\top\theta_* + v^\top G_\gamma^{-1}\Phi^\top\xi + v^\top G_\gamma^{-1}\Phi^\top\eta\ .
\end{aligned}
$$

We study each term separately:

$$
|v^\top G_\gamma^{-1}\Phi^\top\eta| \le \|\eta\|_\infty\|\Phi G_\gamma^{-1}v\|_1
$$

$$\leq h\sqrt{T}\|\Phi G_\gamma^{-1} v\|_2$$
$$\leq h\sqrt{T}\|v\|_{G_\gamma^{-1}} ,$$

with probability at least $1-\delta$

$$|v^\top G_\gamma^{-1}\Phi^\top \xi| \leq \|v\|_{G_\gamma^{-1}} \sqrt{2\sigma^2 \log\left(\frac{2}{\delta}\right)} ,$$

and

$$v^\top G_\gamma^{-1}\Phi^\top \Phi\theta_* - v^\top \theta_* = -\gamma v^\top (\Phi^\top\Phi + \gamma I)^{-1}\theta_*$$

is bounded using Cauchy-Schwarz inequality:

$$\begin{aligned}
&|\gamma v(\Phi^\top\Phi + \gamma I)^{-1}\theta_*| \\
&\leq \gamma\|\theta_*\|\sqrt{v^\top(\Phi^\top\Phi+\gamma I)^{-1} I (\Phi^\top\Phi+\gamma I)^{-1} v} \\
&= \gamma\|\theta_*\|\sqrt{v^\top(\Phi^\top\Phi+\gamma I)^{-1}\gamma^{-1}\gamma I (\Phi^\top\Phi+\gamma I)^{-1} v} \\
&\leq \gamma^{1/2}\|\theta_*\|\sqrt{v^\top(\Phi^\top\Phi+\gamma I)^{-1}(\gamma I + \Phi^\top\Phi)(\Phi^\top\Phi+\gamma I)^{-1} v} \\
&= \gamma^{1/2}\|\theta_*\|\|v\|_{\left(\Phi^\top\Phi+\gamma I\right)^{-1}} \\
&= \gamma^{1/2}\|\theta_*\|\|v\|_{G_\gamma^{-1}} .
\end{aligned}$$

So

$$\begin{aligned}
|v^\top(\widehat{\theta}-\theta_*)| &\leq \sqrt{\gamma}\|\theta_*\|_2\|v\|_{G_\gamma^{-1}} + \|v\|_{G_\gamma^{-1}}\sqrt{2\sigma^2\log\left(\frac{2}{\delta}\right)} + h\sqrt{n}\|v\|_{G_\gamma^{-1}} \\
&= \|v\|_{G_\gamma^{-1}}\left(\sqrt{\gamma}\|\theta_*\|_2 + h\sqrt{T} + \sqrt{2\sigma^2\log\left(\frac{2}{\delta}\right)}\right).
\end{aligned}$$

Union bounding over all $v \in \mathcal{V}$ completes the proof. □

A.6.2 Regret bound

For the same reason as in section A.4, we can consider without loss of generality that ϕ is the identity map in this section. Indeed, the features of the actions - thus denoted x here and $\phi(x)$ in the rest of the paper - appear in this proof only through scalar products.

Theorem 17. *With probability at least $1-\delta$, the regret of Algorithm A.1 satisfies*

$$\sum_{t=1}^{T} \mu_x - \mu_{x_t} \leq c\left(\widetilde{d}(\gamma) + \sqrt{\max_{\mathcal{V}\subset\mathcal{X}} f(\mathcal{V},\gamma)}\left(T(\sqrt{\gamma}\|\theta_*\|_2 + h) + \sqrt{(\sigma^2\log(|\mathcal{X}|\log(T)/\delta))\,T}\right)\right)$$

where $f(\mathcal{V},\gamma) = \inf_{\lambda\in\triangle_\mathcal{V}}\sup_{y\in\mathcal{V}}\|y\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x x x^\top+\gamma I)^{-1}}$ and $\widetilde{d}(\gamma) = \max_{\ell\leq L}\widetilde{d}(\gamma,\lambda_\ell) \leq \max_{\mathcal{V}\subset\mathcal{X}}\sup_{\lambda\in\triangle_\mathcal{V}}\widetilde{d}(\gamma,\lambda)$.

Algorithm A.1. PTR for Regret minimization

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ ($|\mathcal{X}| = n$), feature map ϕ, confidence level $\delta \in (0,1)$, regularization γ, sub-gaussian parameter σ.

Set $\mathcal{X}_1 \leftarrow \mathcal{X}, \ell \leftarrow 1$

while $|\mathcal{X}_\ell| > 1$ **do**

Let $\lambda_\ell \in \triangle_{\mathcal{X}_\ell}$ be a minimizer of $f(\lambda, \mathcal{X}_\ell, \gamma)$ where

$$f(\mathcal{V};\gamma) = \inf_{\lambda \in \triangle_{\mathcal{V}}} f(\lambda, \mathcal{V}, \gamma) = \inf_{\lambda \in \triangle_{\mathcal{V}}} \max_{v \in \mathcal{V}} \|\phi(v)\|^2_{(\sum_{y\in\mathcal{V}} \lambda_y \phi(y)\phi(y)^\top + \gamma I)^{-1}}$$

Set $\epsilon_\ell = 2^{-\ell}$ and $\tau_\ell := \lceil \max\{2\sigma^2 \epsilon_\ell^{-2} f(\mathcal{X}_\ell;\gamma)\log(4\ell^2|\mathcal{X}|/\delta), \widetilde{d}(\gamma,\lambda_\ell)\}\rceil$
Use the PTR procedure of section 2.2.4 to find sparse allocation $\{\widetilde{x}_i\}_{i=1}^{\tau_\ell} \subset \mathcal{X}_\ell$ from λ_ℓ.
Take each action $x \in \{\widetilde{x}_i\}_{i=1}^{\tau_\ell}$ with corresponding features Φ and rewards Y

Compute $\widehat{\theta}_\ell = (\Phi^\top\Phi + \tau_\ell\gamma I)^{-1}\Phi^\top Y$
Update active set:

$$\mathcal{X}_{\ell+1} = \left\{ x \in \mathcal{X}_\ell, \max_{x' \in \mathcal{X}_\ell} \langle \phi(x') - \phi(x), \widehat{\theta}_\ell \rangle < 8\epsilon_\ell \right\}$$

$\ell \leftarrow \ell + 1$

Play unique element of $\mathcal{X}_\ell$ indefinitely.

Recall the definition of $f(\mathcal{V};\gamma) = \inf_{\lambda\in\triangle_\mathcal{V}} \max_{v\in\mathcal{V}} \|v\|^2_{(\sum_{y\in\mathcal{V}}\lambda_y yy^\top + \gamma I)^{-1}}$ and $\bar{f}(\mathcal{X};\gamma) := \max_{\mathcal{V}\subseteq\mathcal{X}} f(\mathcal{V};\gamma)$. Define the event

$$\mathcal{E} := \bigcap_{\ell=1}^{\infty} \bigcap_{x\in\mathcal{X}_\ell} \left\{ |\langle x, \widehat{\theta}_\ell - \theta_* \rangle| \le 2\epsilon_\ell + 2\left(\sqrt{\gamma}\|\theta_*\|_2 + h\right)\sqrt{\bar{f}(\mathcal{X};\gamma)} \right\}$$

Lemma 12. *We have* $\mathbb{P}(\mathcal{E}) \ge 1 - \delta$.

Proof. For any $\mathcal{V} \subseteq \mathcal{X}$ and $x \in \mathcal{V}$ define

$$\mathcal{E}_{\ell,x} = \left\{ |x^\top(\widehat{\theta}_\ell(\mathcal{V}) - \theta_*)| \le 2\epsilon_\ell + 2\left(\sqrt{\gamma}\|\theta_*\|_2 + h\right)\sqrt{\bar{f}(\mathcal{X};\gamma)} \right\}$$

where $\widehat{\theta}_\ell(\mathcal{V})$ is the estimator that would be constructed by the algorithm at stage ℓ with $\mathcal{X}_\ell = \mathcal{V}$. For fixed $\mathcal{V} \subset \mathcal{X}$, $\ell \in \mathbb{N}$, τ_ℓ actions are taken. Thus we apply Lemma 11 with $\tau = \tau_\ell$ and with regularization factor $\tau_\ell\gamma$, so that with probability at least $1 - \frac{\delta}{2\ell^2|\mathcal{X}|}$ we have for any $x \in \mathcal{V}$

$$\begin{aligned}
|x^\top(\widehat{\theta}_\ell - \theta_*)| &\le \|x\|_{\left(\sum_{i=1}^{\tau_\ell}\widetilde{x}_i\widetilde{x}_i^\top + \tau_\ell\gamma I\right)^{-1}} \left(\sqrt{\tau_\ell\gamma}\|\theta_*\|_2 + h\sqrt{\tau_\ell} + \sqrt{2\sigma^2\log(4\ell^2|\mathcal{X}|/\delta)}\right) \\
&\le 2\|x\|_{(\tau_\ell A(\lambda_\ell) + \tau_\ell\gamma I)^{-1}} \left(\sqrt{\tau_\ell\gamma}\|\theta_*\|_2 + h\sqrt{\tau_\ell} + \sqrt{2\sigma^2\log(4\ell^2|\mathcal{X}|/\delta)}\right) \\
&= 2\|x\|_{(A(\lambda_\ell)+\gamma I)^{-1}} \left(\sqrt{\gamma}\|\theta_*\|_2 + h + \sqrt{\frac{2\sigma^2\log(4\ell^2|\mathcal{X}|/\delta)}{\tau_\ell}}\right)
\end{aligned}$$

$$\leq 2\sqrt{f(\mathcal{V};\gamma)}\left(\sqrt{\gamma}\|\theta_*\|_2 + h + \epsilon_\ell/\sqrt{f(\mathcal{V};\gamma)}\right)$$
$$\leq 2\epsilon_\ell + 2\sqrt{\bar{f}(\mathcal{X};\gamma)}\,(\sqrt{\gamma}\|\theta_*\|_2 + h)$$

Noting that $\mathcal{E} := \bigcap_{\ell=1}^{\infty}\bigcap_{x\in\mathcal{X}_\ell}\mathcal{E}_{x,\ell}(\mathcal{X}_\ell)$, the rest of the proof with the robust estimator applies here. □

The next lemma is similar to the one for the robust estimator, and the proof will follow the same argument as for the robust estimator.

Lemma 13. *For all $\ell \in \mathbb{N}$ we have* $\max_{x\in\mathcal{X}_\ell}\mu_* - \mu_x \leq \max\{32\epsilon_\ell, 32(4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\}$.

Proof. An arm $x \in \mathcal{X}_\ell$ is discarded (i.e., not in $\mathcal{X}_{\ell+1}$) if $\max_{x'\in\mathcal{X}_\ell}\langle x', \widehat{\theta}\rangle - \langle x, \widehat{\theta}\rangle > 8\epsilon_\ell$. Let $\bar{\ell} := \max\{\ell : \epsilon_\ell > (4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\}$. If $x_* = \arg\max_{x\in\mathcal{X}}\mu_x$ then $x_* \in \mathcal{X}_1$. Now if $x_* \in \mathcal{X}_\ell$ for some $\ell \leq \bar{\ell}$, then for any $x' \in \mathcal{X}_\ell$ we have

$$\begin{aligned}
\langle x', \widehat{\theta}\rangle - \langle x_*, \widehat{\theta}\rangle &\leq \langle x' - x_*, \theta_*\rangle + 4\epsilon_\ell + 4(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\leq \mu_x - \mu_{x_*} + 2h + 4\epsilon_\ell + 4(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\leq 4\epsilon_\ell + (4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\leq 8\epsilon_\ell
\end{aligned}$$

which implies $x_* \in \mathcal{X}_{\ell+1}$. Moreover, suppose that $\ell \leq \bar{\ell}$ and there exists some $x \in \mathcal{X}_\ell$ such that $\mu_* - \mu_x > 16\epsilon_\ell$, then

$$\begin{aligned}
\max_{x'\in\mathcal{X}_\ell}\langle x', \widehat{\theta}\rangle - \langle x, \widehat{\theta}\rangle &\geq \langle x_*, \widehat{\theta}\rangle - \langle x, \widehat{\theta}\rangle \\
&\geq \langle x_* - x, \theta_*\rangle - 4\epsilon_\ell - 4(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\geq \mu_* - \mu_x - 2h - 4\epsilon_\ell - 4(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\geq \mu_* - \mu_x - 4\epsilon_\ell - (4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)} \\
&\geq \mu_* - \mu_x - 8\epsilon_\ell \\
&> 8\epsilon_\ell
\end{aligned}$$

which implies $\max_{x\in\mathcal{X}_{\ell+1}}\mu_* - \mu_x \leq 16\epsilon_\ell = 32\epsilon_{\ell+1}$. Because $\mathcal{X}_{\ell+1} \subseteq \mathcal{X}_\ell$ we have for $\ell > \bar{\ell}$ that

$$\begin{aligned}
\max_{x\in\mathcal{X}_\ell}\mu_* - \mu_x &\leq \max_{x\in\mathcal{X}_{\bar{\ell}+1}}\mu_* - \mu_x \\
&\leq 32\epsilon_{\bar{\ell}+1} \\
&\leq 32(4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)}.
\end{aligned}$$

Thus, $\max_{x\in\mathcal{X}_\ell}\mu_* - \mu_x \leq \max\{32\epsilon_\ell, 32(4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\}$. □

We now compute the final regret bound. After T steps of the algorithm, let T_x denote the number of times arm x is played. Let $\Gamma = (4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)}$. If L is the final round reached after T steps, we have

$$\sum_{x\in\mathcal{X}}(\mu_* - \mu_x)T_x \leq \sum_{\ell=1}^{L}\max_{x\in\mathcal{X}_\ell}(\mu_* - \mu_x)\tau_\ell$$

$$
\begin{aligned}
&\leq \sum_{\ell=1}^{L} \tau_\ell \max\{32\epsilon_\ell, 32(4\sqrt{\gamma}\|\theta_*\|_2 + 6h)\sqrt{\bar{f}(\mathcal{X};\gamma)}\} \\
&\leq \sum_{\ell=1}^{L} \tau_\ell \max\{32\epsilon_\ell, 32\Gamma\} \\
&\leq \sum_{\ell:\epsilon_\ell<\Gamma} 32\Gamma\tau_\ell + 32 \sum_{\ell:\epsilon_\ell\geq\Gamma} \epsilon_\ell\tau_\ell \\
&\leq \sum_{\ell:\epsilon_\ell<\Gamma} 32\Gamma\tau_\ell + 32\nu T + \sum_{\ell:\epsilon_\ell\geq\Gamma\vee\nu} 32\epsilon_\ell\tau_\ell \\
&\leq \sum_{\ell:\epsilon_\ell<\Gamma} 32\Gamma\tau_\ell + 32\nu T + \sum_{\ell:\epsilon_\ell\geq\nu} 32\epsilon_\ell\tau_\ell \\
&\leq c\left(\Gamma T + \nu T + \sum_{\ell:\epsilon_\ell\geq\nu} \epsilon_\ell \cdot \left(2\sigma^2\epsilon_\ell^{-2} f(\mathcal{X}_\ell;\gamma)\log(4\ell^2|\mathcal{X}|/\delta) + \widetilde{d}(\gamma,\lambda_\ell)\right)\right) \\
&\leq c\left(\Gamma T + \nu T + \left(2\sigma^2\bar{f}(\mathcal{X};\gamma)\log(4\lceil\log_2(1/\nu)\rceil^2|\mathcal{X}|/\delta) + \widetilde{d}(\gamma)\right)\sum_{\ell:\epsilon_\ell\geq\nu}\epsilon_\ell^{-1}\right) \\
&\leq c\left(\Gamma T + \nu T + \nu^{-1}\left(2\sigma^2\bar{f}(\mathcal{X};\gamma)\log(4\lceil\log_2(1/\nu)\rceil^2|\mathcal{X}|/\delta)\right) + 32\widetilde{d}(\gamma)\right).
\end{aligned}
$$

Where we denote $\widetilde{d}(\gamma) = \max_{\ell\leq L}\widetilde{d}(\gamma,\lambda_\ell)$. Choosing $\nu = \sqrt{\left(2\sigma^2\bar{f}(\mathcal{X};\gamma)\log(|\mathcal{X}|/\delta)\right)/T}$ and plugging Γ back in yields

$$
\sum_{x\in\mathcal{X}}(\mu_* - \mu_x)T_x \leq c\left(\widetilde{d}(\gamma) + \sqrt{\bar{f}(\mathcal{X};\gamma)}\left(T(\sqrt{\gamma}\|\theta_*\|_2 + h) + \sqrt{(\sigma^2\log(|\mathcal{X}|\log(T)/\delta))\,T}\right)\right).
$$

Choosing $\gamma = 1/T$ yields

$$
\sum_{x\in\mathcal{X}}(\mu_* - \mu_x)T_x \leq c\left(\widetilde{d}(\gamma) + \sqrt{\bar{f}(\mathcal{X};1/T)}\left(hT + \sqrt{((\|\theta_*\|^2 + \sigma^2)\log(|\mathcal{X}|\log(T)/\delta))\,T}\right)\right).
$$

A.6.3 Sample complexity bound

For any $\mathcal{V}\subset\mathcal{Z}$ define $f(\mathcal{X},\mathcal{V};\gamma) = \min_{\lambda\in\triangle_\mathcal{X}}\max_{z,z'\in\mathcal{V}}\|\phi(z)-\phi(z')\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I)^{-1}}$

Theorem 18. *With $z_* \in \arg\max_{z\in\mathcal{Z}}\langle z,\theta_*\rangle$, fix any $\epsilon\geq\bar{\epsilon}$ where*

$$
\bar{\epsilon} = 8\min\{\epsilon\geq 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X},\{z\in\mathcal{Z} : \langle\phi(z_*)-\phi(z),\theta_*\rangle\leq\epsilon\};\gamma)}) \leq \epsilon\}.
$$

Then with probability at least $1-\delta$, once the algorithm has taken at least τ samples where

$$
\tau \leq c'\left(\widetilde{d}(\gamma)\log_2(8(\Delta\vee\bar{\epsilon})^{-1}) + \lceil\log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil\sigma^2\log(2\lceil\log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil^2|\mathcal{Z}|^2/\delta)\rho^*(\gamma,\bar{\epsilon})\right)
$$

we have that $\mu_{\widehat{z}} \geq \max_{z'\in\mathcal{Z}} -\epsilon$ where $\widehat{z}$ is any arm in the set $\mathcal{Z}_\ell$ under consideration after τ pulls and

$$
\rho^*(\gamma,\epsilon) = \inf_{\lambda\in\Delta_\mathcal{X}}\sup_{z\in\mathcal{Z}}\frac{\|\phi(z_*)-\phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{\max\{\epsilon^2,\langle\theta_*,\phi(z_*)-\phi(z)\rangle^2)\}} \tag{A.1}
$$

Algorithm A.2. PTR for Pure exploration

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Z} \subset \mathbb{R}^d$, feature map ϕ, confidence level $\delta \in (0,1)$, regularization γ, sub-gaussian parameter σ, norm of model parameter B, bound on the misspecification noise h.

Let $\mathcal{Z}_1 \leftarrow \mathcal{Z}, \ell \leftarrow 1$

while $|\mathcal{Z}_\ell| > 1$ **do**

Let $\lambda_\ell \in \triangle_\mathcal{X}$ be a minimizer of $f(\lambda, \mathcal{Z}_\ell, \gamma)$ where

$$f(\mathcal{V};\gamma) = \inf_{\lambda\in\triangle_\mathcal{X}} f(\lambda,\mathcal{V},\gamma) = \inf_{\lambda\in\triangle_\mathcal{X}} \max_{v,v'\in\mathcal{V}} \|\phi(v)-\phi(v')\|^2_{(\sum_{x\in\mathcal{X}}\lambda_y\phi(x)\phi(x)^\top+\gamma I)^{-1}}$$

Set $\epsilon_\ell = 2^{-\ell}$ and $\tau_\ell := \lceil \max\{2\sigma^2\epsilon_\ell^{-2} f(\mathcal{Z}_\ell;\gamma)\log(4\ell^2|\mathcal{Z}|/\delta), \widetilde{d}(\gamma,\lambda_\ell)\}\rceil$
Use the PTR procedure of section 2.2.4 to find sparse allocation $\{\widetilde{x}_i\}_{i=1}^{\tau_\ell} \subset \mathcal{X}_\ell$ from λ_ℓ.
Take each action $x \in \{\widetilde{x}_i\}_{i=1}^{\tau_\ell}$ with corresponding features Φ and rewards Y

Compute $\widehat{\theta}_\ell = (\Phi^\top\Phi + \tau_\ell\gamma I)^{-1}\Phi^\top Y$

$$\mathcal{Z}_{\ell+1} \leftarrow \mathcal{Z}_\ell \setminus \left\{z \in \mathcal{Z}_\ell : \max_{z'\in\mathcal{Z}_\ell} \langle \phi(z')-\phi(z), \widehat{\theta}_\ell\rangle > \epsilon_\ell\right\}$$

$\ell \leftarrow \ell + 1$

Output: $\mathcal{Z}_\ell$

and $\widetilde{d}(\gamma) = \max_{\ell\le\lceil\log_2(8(\Delta\vee\epsilon)^{-1})\rceil} \widetilde{d}(\gamma,\lambda_\ell) \le \max_{\mathcal{V}\subset\mathcal{X}} \sup_{\lambda\in\triangle_\mathcal{V}} \widetilde{d}(\gamma,\lambda)$.

We first prove the following intermediate result.

Theorem 19. *Recall that we defined*

$$\bar{\epsilon} = 8\min\{\epsilon \ge 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z\in\mathcal{Z} : \langle\phi(z_*)-\phi(z),\theta_*\rangle \le \epsilon\};\gamma)}) \le \epsilon\}.$$

Then $\max_{z\in\mathcal{Z}_\ell} \mu_* - \mu_z \le 16\max\{\epsilon_\ell, \bar{\epsilon}\}$ *for all* $\ell \ge 0$ *with probability at least* $1-\delta$.

Define the event

$$\mathcal{E} := \bigcap_{\ell=1}^\infty \bigcap_{z,z'\in\mathcal{Z}_\ell} \left\{|\langle\phi(z)-\phi(z'), \widehat{\theta}_\ell - \theta_*\rangle| \le 2\epsilon_\ell + 2\left(\sqrt{\gamma}\|\theta_*\|_2 + h\right)\sqrt{f(\mathcal{X},\mathcal{Z}_\ell;\gamma)}\right\}$$

Lemma 14. *We have* $\mathbb{P}(\mathcal{E}) \ge 1-\delta$.

Proof. For any $\mathcal{V} \subseteq \mathcal{Z}$ and $x \in$ define

$$\mathcal{E}_{z,z',\ell}(\mathcal{V}) = \left\{|\langle\phi(z)-\phi(z'), \widehat{\theta}_\ell(\mathcal{V}) - \theta_*\rangle| \le 2\epsilon_\ell + 2\left(\sqrt{\gamma}\|\theta_*\|_2 + h\right)\sqrt{f(\mathcal{X},\mathcal{V};\gamma)}\right\}$$

where $\widehat{\theta}_\ell(\mathcal{V})$ is the estimator that would be constructed by the algorithm at stage ℓ with $\mathcal{Z}_\ell = \mathcal{V}$. For fixed $\mathcal{V} \subset \mathcal{Z}$, $\ell \in \mathbb{N}$ and $x \in \mathcal{V}$, τ_ℓ actions are taken. Thus we apply Lemma 11 with $\tau = \tau_\ell$ and with regularization factor $\tau_\ell\gamma$, so that with probability at least $1 - \frac{\delta}{2\ell^2|\mathcal{Z}|}$ we have for any $z, z' \in \mathcal{V}$

$$|(\phi(z)-\phi(z'))^\top(\widehat{\theta}_\ell - \theta_*)|$$

$$
\begin{aligned}
&\le 2\|\phi(z)-\phi(z')\|_{\left(\sum_{i=1}^{\tau_\ell}\phi(\widetilde{x}_i)\phi(\widetilde{x}_i)^\top+\tau_\ell\gamma I\right)^{-1}}\left(\sqrt{\tau_\ell\gamma}\|\theta_*\|_2+h\sqrt{\tau_\ell}+\sqrt{2\sigma^2\log(4\ell^2|\mathcal{Z}|/\delta)}\right)\\
&\le 2\|\phi(z)-\phi(z')\|_{(\tau_\ell A(\lambda)+\tau_\ell\gamma I)^{-1}}\left(\sqrt{\tau_\ell\gamma}\|\theta_*\|_2+h\sqrt{\tau_\ell}+\sqrt{2\sigma^2\log(4\ell^2|\mathcal{Z}|/\delta)}\right)\\
&= 2\|\phi(z)-\phi(z')\|_{(A(\lambda)+\gamma I)^{-1}}\left(\sqrt{\gamma}\|\theta_*\|_2+h+\sqrt{\frac{2\sigma^2\log(4\ell^2|\mathcal{Z}|/\delta)}{\tau_\ell}}\right)\\
&\le 2\sqrt{f(\mathcal{X},\mathcal{V};\gamma)}\left(\sqrt{\gamma}\|\theta_*\|_2+h+\epsilon_\ell/\sqrt{f(\mathcal{X},\mathcal{V};\gamma)}\right)\\
&= 2\epsilon_\ell+2\sqrt{f(\mathcal{X},\mathcal{V};\gamma)}\left(\sqrt{\gamma}\|\theta_*\|_2+h\right)
\end{aligned}
$$

Noting that $\mathcal{E}:=\bigcap_{\ell=1}^{\infty}\bigcap_{z,z'\in\mathcal{Z}_\ell}\mathcal{E}_{z,z',\ell}(\mathcal{Z}_\ell)$, the rest of the proof with the robust estimator applies here. □

Lemma 15. *For all $\ell\in\mathbb{N}$ we have* $\max_{z\in\mathcal{Z}_\ell}\mu_*-\mu_z\le 16\max\{\epsilon_\ell,\epsilon_{\bar{\ell}}\}$.

Proof. An arm $z\in\mathcal{Z}_\ell$ is discarded (i.e., not in $\mathcal{Z}_{\ell+1}$) if $\max_{z'\in\mathcal{Z}_\ell}\langle\phi(z')-\phi(z),\widehat{\theta}\rangle>4\epsilon_\ell$.

Define $S_1=\mathcal{Z}$ and $S_{\ell+1}=\{z\in S_\ell:\langle\phi(z_*)-\phi(z),\theta_*\rangle\le 6\epsilon_\ell+2(\sqrt{\gamma}\|\theta_*\|_2+h)\sqrt{f(\mathcal{X},S_\ell;\gamma)}\}$. Define

$$
\bar{\ell}=\max\{\ell:(\sqrt{\gamma}\|\theta_*\|_2+h)(2+\sqrt{f(\mathcal{X},S_\ell;\gamma)})\le\epsilon_\ell\}.
$$

We will show $\{z_*\in\mathcal{Z}_\ell\}\cap\{\mathcal{Z}_\ell\subset S_\ell\}\cap\{\ell\le\bar{\ell}\}\implies\{z_*\in\mathcal{Z}_{\ell+1}\}\cap\{\mathcal{Z}_{\ell+1}\subset S_{\ell+1}\}$. Noting that $\{z_*\in\mathcal{Z}_\ell\}\cap\{\mathcal{Z}_\ell\subset S_\ell\}$ holds for $\ell=1$, we will assume an inductive hypothesis of this condition for some $\ell\le\bar{\ell}$.

First we will show $\{z_*\in\mathcal{Z}_\ell\}\cap\{\mathcal{Z}_\ell\subset S_\ell\}\cap\{\ell\le\bar{\ell}\}\implies\{z_*\in\mathcal{Z}_{\ell+1}\}$. On $\{z_*\in\mathcal{Z}_\ell\}\cap\{\mathcal{Z}_\ell\subset S_\ell\}\cap\{\ell\le\bar{\ell}\}$, we have for any $z'\in\mathcal{Z}_\ell$ that

$$
\begin{aligned}
\langle\phi(z')-\phi(z_*),\widehat{\theta}\rangle&\le\langle\phi(z')-\phi(z_*),\theta_*\rangle+2\epsilon_\ell+2(\sqrt{\gamma}\|\theta_*\|_2+h)\sqrt{f(\mathcal{X},\mathcal{Z}_\ell;\gamma)}\\
&\le\mu_{z'}-\mu_{z_*}+2h+2\epsilon_\ell+2(\sqrt{\gamma}\|\theta_*\|_2+h)\sqrt{f(\mathcal{X},\mathcal{Z}_\ell;\gamma)}\\
&\le 2\epsilon_\ell+2(\sqrt{\gamma}\|\theta_*\|_2+h)(2+\sqrt{f(\mathcal{X},\mathcal{Z}_\ell;\gamma)})\\
&\le 2\epsilon_\ell+2(\sqrt{\gamma}\|\theta_*\|_2+h)(2+\sqrt{f(\mathcal{X},S_\ell;\gamma)})\\
&\le 4\epsilon_\ell
\end{aligned}
$$

which implies z_* is not eliminated, that is, $z_*\in\mathcal{Z}_{\ell+1}$. The second-to-last inequality follows from

$$
\begin{aligned}
f(\mathcal{X},\mathcal{Z}_\ell;\gamma)&=\inf_{\lambda}\max_{z,z'\in\mathcal{Z}_\ell}\|\phi(z)-\phi(z')\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I)^{-1}}\\
&\le\inf_{\lambda}\max_{z,z'\in S_\ell}\|\phi(z)-\phi(z')\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I)^{-1}}\\
&=f(\mathcal{X},S_\ell;\gamma).
\end{aligned}
$$

Now we will show $\{z_*\in\mathcal{Z}_\ell\}\cap\{\mathcal{Z}_\ell\subset S_\ell\}\cap\{\ell\le\bar{\ell}\}\implies\{\mathcal{Z}_{\ell+1}\subset S_{\ell+1}\}$. For any $z\in\mathcal{Z}_\ell\cap S_{\ell+1}^c$ we have

$$
\begin{aligned}
\max_{z'\in\mathcal{Z}_\ell}\langle\phi(z')-\phi(z),\widehat{\theta}\rangle&\ge\langle\phi(z_*)-\phi(z),\widehat{\theta}\rangle\\
&\ge\langle\phi(z_*)-\phi(z),\theta_*\rangle-2\epsilon_\ell-2(\sqrt{\gamma}\|\theta_*\|_2+h)\sqrt{f(\mathcal{X},\mathcal{Z}_\ell;\gamma)}\\
&>6\epsilon_\ell+2(\sqrt{\gamma}\|\theta_*\|_2+h)\sqrt{f(\mathcal{X},S_\ell;\gamma)}-2\epsilon_\ell-2(\sqrt{\gamma}\|\theta_*\|_2+h)\sqrt{f(\mathcal{X},\mathcal{Z}_\ell;\gamma)}
\end{aligned}
$$

$$\geq 4\epsilon_\ell$$

which implies $z \notin \mathcal{Z}_{\ell+1}$, and $\mathcal{Z}_{\ell+1} \subset S_{\ell+1}$.

Thus, for $\ell \leq \bar{\ell}$ we have

$$\begin{aligned}
\max_{z\in\mathcal{Z}_\ell} \mu_* - \mu_z &\leq \max_{z\in\mathcal{Z}_\ell} \langle \phi(z_*) - \phi(z), \theta_* \rangle + 2h \\
&\leq \max_{z\in\mathcal{S}_\ell} \langle \phi(z_*) - \phi(z), \theta_* \rangle + 2h \\
&\leq 6\epsilon_{\ell-1} + 2h + 2(\sqrt{\gamma}\|\theta_*\|_2 + h)\sqrt{f(\mathcal{X}, S_{\ell-1}; \gamma)} \\
&\leq 6\epsilon_{\ell-1} + 2(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, S_{\ell-1}; \gamma)}) \\
&\leq 8\epsilon_{\ell-1} = 16\epsilon_\ell.
\end{aligned}$$

And because $\mathcal{Z}_{\ell+1} \subseteq \mathcal{Z}_\ell$ we always have that $\max_{z\in\mathcal{Z}_\ell} \mu_* - \mu_z \leq 16 \max\{\epsilon_\ell, \epsilon_{\bar{\ell}}\}$. Note that

$$\begin{aligned}
\bar{\ell} &= \max\{\ell : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, S_\ell; \gamma)}) \leq \epsilon_\ell\} \\
&\geq \max\{\ell : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \leq 4\epsilon_\ell\}; \gamma)}) \leq \epsilon_\ell\} \\
&= \max\{\ell : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \leq 4\epsilon_\ell\}; \gamma)}) \leq 4\epsilon_\ell\} \\
&= -2 + \max\{\ell : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \leq \epsilon_\ell\}; \gamma)}) \leq \epsilon_\ell\} \\
&\geq -3 - \log_2(\min\{\epsilon > 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{z \in \mathcal{Z} : \langle \phi(z_*) - \phi(z), \theta_* \rangle \leq \epsilon\}; \gamma)}) \leq \epsilon\})
\end{aligned}$$

which defines $\bar{\epsilon}$. □

Denoting $\widetilde{d}(\gamma) = \max_{\ell \leq \lceil \log_2(8(\Delta \vee \bar{\epsilon})^{-1})\rceil} \widetilde{d}(\gamma, \lambda_\ell)$, the sample complexity to return an $\bar{\epsilon}$-good arm is equal to

$$\begin{aligned}
&\sum_{\ell=1}^{\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil} \tau_\ell \\
&\leq \sum_{\ell=1}^{\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil} (2\epsilon_\ell^{-2}(\gamma) f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)\sigma^2 \log(2\ell^2|\mathcal{Z}|^2/\delta) + \widetilde{d}(\gamma, \lambda_\ell)) \\
&\leq c\left(\widetilde{d}(\gamma)\log_2(8(\Delta\vee\bar{\epsilon})^{-1}) + \sigma^2 \log(2\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil^2|\mathcal{Z}|^2/\delta) \sum_{\ell=1}^{\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil} \epsilon_\ell^{-2} f(\mathcal{X}, \mathcal{Z}_\ell; \gamma)\right) \\
&\leq c\left(\widetilde{d}(\gamma)\log_2(8(\Delta\vee\bar{\epsilon})^{-1}) + \sigma^2 \log(2\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil^2|\mathcal{Z}|^2/\delta) \sum_{\ell=1}^{\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil} \epsilon_\ell^{-2} f(\mathcal{X}, S_\ell; \gamma)\right) \\
&\leq c'\left(\widetilde{d}(\gamma)\log_2(8(\Delta\vee\bar{\epsilon})^{-1}) + \lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil \sigma^2 \log(2\lceil \log_2(8(\Delta\vee\bar{\epsilon})^{-1})\rceil^2|\mathcal{Z}|^2/\delta)\rho^*(\gamma, \bar{\epsilon})\right)
\end{aligned}$$

where the last line follows from

$$\rho^*(\gamma, \bar{\epsilon}) = \inf_{\lambda\in\Delta_\mathcal{X}} \sup_{z\in\mathcal{Z}} \frac{\|\phi(z_*) - \phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I\right)^{-1}}}{\max\{\bar{\epsilon}^2, \langle \phi(z_*) - \phi(z), \theta_* \rangle^2)\}}$$

$$
\begin{aligned}
&= \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{\ell\le\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sup_{z\in S_\ell} \frac{\|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}}{\max\{\bar\epsilon^2,((\phi(z_*)-\phi(z))^\top\theta_*)^2\}} \\
&\ge \inf_{\lambda\in\Delta_{\mathcal{X}}} \frac{1}{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sup_{z\in S_\ell} \frac{\|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}}{\max\{\bar\epsilon^2,((\phi(z_*)-\phi(z))^\top\theta_*)^2\}} \\
&\ge \frac{1}{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{z\in S_\ell} \frac{\|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}}{\max\{\bar\epsilon^2,((\phi(z_*)-\phi(z))^\top\theta_*)^2\}} \\
&\ge \frac{1}{4\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} 2^{2\ell} \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{z\in S_\ell} \|\phi(z_*)-\phi(z)\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}} \\
&\ge \frac{1}{16\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} 2^{2\ell} \inf_{\lambda\in\Delta_{\mathcal{X}}} \sup_{z,z'\in S_\ell} \|\phi(z)-\phi(z')\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}} \\
&= \frac{1}{16\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} \sum_{\ell=1}^{\lceil\log_2(8(\Delta\vee\bar\epsilon)^{-1})\rceil} 2^{2\ell} f(\mathcal{X},S_\ell;\gamma)\ .
\end{aligned}
$$

A.7 Related work results

Lemma 16. *If* $\lambda^*\in\arg\max_{\lambda\in\triangle_{\mathcal{V}}} f(\lambda)$ *where* $f(\lambda)=\log\left(\det\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)\right)$*, then*

$$
\begin{aligned}
\sup_{x\in\mathcal{X}}\|\phi(x)\|^2_{A^{(\gamma)}(\lambda^*)^{-1}} &= \sum_{x\in\mathcal{X}}\lambda^*_x\|\phi(x)\|^2_{A^{\gamma}(\lambda^*)^{-1}} \\
&= \textit{Trace}\left(A(\lambda^*)(A(\lambda^*)+\gamma I)^{-1}\right) \\
&= \textit{Trace}\left(K_{\lambda^*}(K_{\lambda^*}+\gamma I)^{-1}\right)
\end{aligned}
$$

Proof. We first state that

$$
\begin{aligned}
\sum_{x\in\mathcal{X}}\lambda^*_x\|\phi(x)\|^2_{A^{\gamma}(\lambda^*)^{-1}} &= \sum_{x\in\mathcal{X}}\lambda^*_x\phi(x)^\top A^{\gamma}(\lambda^*)^{-1}\phi(x) \\
&= \text{Trace}\left(\sum_{x\in\mathcal{X}}\lambda^*_x\phi(x)^\top A^{\gamma}(\lambda^*)^{-1}\phi(x)\right) \\
&= \text{Trace}\left(\sum_{x\in\mathcal{X}}\lambda^*_x\phi(x)\phi(x)^\top A^{\gamma}(\lambda^*)^{-1}\right) \\
&= \text{Trace}\left(A(\lambda^*)(A(\lambda^*)+\gamma I)^{-1}\right)
\end{aligned}
$$

This implies

$$
\sup_{x\in\mathcal{X}}\|\phi(x)\|^2_{A^{(\gamma)}(\lambda^*)^{-1}} \ge \sum_{x\in\mathcal{X}}\lambda^*_x\|\phi(x)\|^2_{A^{\gamma}(\lambda^*)^{-1}} = \text{Trace}\left(A(\lambda^*)(A(\lambda^*)+\gamma I)^{-1}\right)
$$

Further, one can compute that

$$
[\nabla_\lambda f(\lambda^*)]_x = \text{Trace}\left(A^{(\gamma)}(\lambda^*)^{-1}xx^\top\right) = \|\phi(x)\|^2_{A^{(\gamma)}(\lambda^*)^{-1}}
$$

And last, λ^* satisfies the first order conditions on f: for any $\lambda \in \triangle_{\mathcal{X}}$

$$\begin{aligned}
0 &\geq \langle \nabla_\lambda f(\lambda^*), \lambda - \lambda^* \rangle \\
&= \sum_{x \in \mathcal{X}} \lambda_x \|\phi(x)\|^2_{A^\gamma(\lambda^*)^{-1}} - \sum_{x \in \mathcal{X}} \lambda^*_x \|\phi(x)\|^2_{A^\gamma(\lambda^*)^{-1}} \\
&= \sum_{x \in \mathcal{X}} \lambda_x \|\phi(x)\|^2_{A^\gamma(\lambda^*)^{-1}} - \text{Trace}\left(A(\lambda^*)(A(\lambda^*) + \gamma I)^{-1}\right)
\end{aligned}$$

Choosing λ to be a Dirac at $\arg\max_{x \in \mathcal{X}} \|\phi(x)\|^2_{A^\gamma(\lambda^*)^{-1}}$, we get to

$$\max_{x \in \mathcal{X}} \|\phi(x)\|^2_{A^\gamma(\lambda^*)^{-1}} \leq \text{Trace}\left(A(\lambda^*)(A(\lambda^*) + \gamma I)^{-1}\right).$$

Hence the result of the lemma. □

Lemma 17. *We can lower bound γ_T the notion of information gain from (Srinivas et al., 2009) as*

$$\gamma_T \geq \frac{2}{3} \max_{\mathcal{V} \subset \mathcal{X}} \inf_{\lambda \in \triangle_{\mathcal{V}}} \sup_{x \in \mathcal{V}} \|\phi(x)\|^2_{\left(\sum_{x \in \mathcal{V}} \lambda_x \phi(x)\phi(x)^\top + \gamma/T I\right)^{-1}} + |\mathcal{X}| \log(\gamma) \ .$$

Proof. Recall the definition of (Srinivas et al., 2009) notion of information gain:

$$\gamma_T := \sup_{\lambda \in \triangle_{\mathcal{X}}} \log(\det(TK_\lambda + \gamma I))$$

where K_λ is defined in Section 2.2.3. Note that the case where we have an infinite dimensional RKHS and ϕ is any feature map reduces to the finite one with ϕ being the identity map by computing $\Phi_\lambda \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ such that $K_\lambda = \Phi_\lambda \Phi_\lambda^\top$ and then looking at the (finite dimension) columns of Φ_λ. So we can write without loss of generality

$$\gamma_T = \sup_{\lambda \in \triangle_{\mathcal{X}}} \log\left(\det\left(T \sum_{x \in \mathcal{X}} \lambda_x x x^\top + \gamma I\right)\right)$$

Thus

$$\gamma_T = \sup_{\lambda \in \triangle_{\mathcal{X}}} \log\left(\det\left(T \sum_{x \in \mathcal{X}} \lambda_x x x^\top + \gamma I\right)\right) \geq \sup_{\mathcal{V} \subset \mathcal{X}} \sup_{\lambda \in \triangle_{\mathcal{V}}} \log\left(\det\left(T \sum_{x \in \mathcal{V}} \lambda_x x x^\top + \gamma I\right)\right)$$

Fix for now $\mathcal{V} \subset \mathcal{X}$ and let $\lambda^* \in \triangle_{\mathcal{V}}$ be such that

$$\lambda^* \in \arg\max_{\lambda \in \triangle_{\mathcal{V}}} \log\left(\det\left(T \sum_{x \in \mathcal{V}} \lambda_x x x^\top + \gamma I\right)\right) = \arg\max_{\lambda \in \triangle_{\mathcal{V}}} \log\left(\det\left(\sum_{x \in \mathcal{V}} \lambda_x x x^\top + \gamma/T I\right)\right)$$

Inspired from equation 19.9 of (Lattimore and Szepesvári, 2020), we write for some $x_0 \in \mathcal{V}$

$$\begin{aligned}
&\det\left(T \sum_{x \in \mathcal{V}} \lambda^*_x x x^\top + \gamma I\right) \\
&= \det\left(T \sum_{x \in \mathcal{V} \setminus \{x_0\}} \lambda^*_x x x^\top + \gamma I + T\lambda^*_{x_0} x_0 x_0^\top\right)
\end{aligned}$$

$$= \det\left(T\sum_{x\in\mathcal{V}\setminus\{x_0\}}\lambda_x^* xx^\top + \gamma I\right)\det\left(I + \left(T\sum_{x\in\mathcal{V}\setminus\{x_0\}}\lambda_x^* xx^\top + \gamma I\right)^{-1/2} T\lambda_{x_0}^* x_0 x_0^\top \left(T\sum_{x\in\mathcal{V}\setminus\{x_0\}}\lambda_x^* xx^\top + \gamma I\right)^{-1/}\right.$$

$$= \det\left(T\sum_{x\in\mathcal{V}\setminus\{x_0\}}\lambda_x^* xx^\top + \gamma I\right)\left(1 + T\lambda_{x_0}^*\|x_0\|^2_{\left(T\sum_{x\in\mathcal{V}\setminus\{x_0\}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}\right)$$

$$\geq \det\left(T\sum_{x\in\mathcal{V}\setminus\{x_0\}}\lambda_x^* xx^\top + \gamma I\right)\left(1 + T\lambda_{x_0}^*\|x_0\|^2_{\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}\right)$$

We can now iterate with all the remaining $x \in \mathcal{V}\setminus\{x_0\}$, to get

$$\det\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top + \gamma I\right) \geq \det(\gamma I)\prod_{x_0\in\mathcal{V}}\left(1 + T\lambda_{x_0}^*\|x_0\|^2_{\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}\right)$$

equivalent to

$$\log\det\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top + \gamma I\right) \geq \log\det(\gamma I) + \sum_{x\in\mathcal{V}}\log\left(1 + T\lambda_x^*\|x\|^2_{\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}\right)$$

We know that if $x \geq 0$ holds $\log(1+x) \geq 2x/(2+x)$. Note that $\|x\|^2_{\left(\sum_{x'\in\mathcal{V}}\lambda_{x'}x'x'^\top+\gamma/TI\right)^{-1}} \leq 1$ always holds:

$$\|x\|^2_{\left(\sum_{x'\in\mathcal{V}}\lambda_{x'}x'x'^\top+\gamma/TI\right)^{-1}} = \|x\|^2_{\left(\sum_{x'\in\mathcal{V}\setminus\{x\}}\lambda_{x'}x'x'^\top+\gamma/TI+xx^\top\right)^{-1}} = \alpha - \alpha^2/(1+\alpha) = \alpha/(1+\alpha) \leq 1\,.$$

where we used Sherman–Morrison formula and defined $\alpha = \|x\|^2_{\left(\sum_{x'\in\mathcal{V}\setminus\{x\}}\lambda_{x'}x'x'^\top+\gamma/TI\right)^{-1}}$. Thus holds $0 \leq T\lambda_x^*\|x\|^2_{\left(T\sum_{x'\in\mathcal{V}}\lambda_{x'}^* x'x'^\top+\gamma I\right)^{-1}} \leq 1$. So

$$\sum_{x\in\mathcal{V}}\log\left(1 + T\lambda_x^*\|x\|^2_{\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}\right) \geq \frac{2}{2+1}\sum_{x\in\mathcal{V}}T\lambda_x^*\|x\|^2_{\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}$$

And thus

$$\begin{aligned}\frac{3}{2}\log\left(\frac{\det\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top + \gamma I\right)}{\det(\gamma I)}\right) &\geq \sum_{x\in\mathcal{V}}T\lambda_x^*\|x\|^2_{\left(T\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma I\right)^{-1}}\\ &= \sum_{x\in\mathcal{V}}\lambda_x^*\|x\|^2_{\left(\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma/TI\right)^{-1}}\\ &= \sup_{x\in\mathcal{V}}\|x\|^2_{\left(\sum_{x\in\mathcal{V}}\lambda_x^* xx^\top+\gamma/TI\right)^{-1}}\,.\end{aligned}$$

Where the last equality comes from lemma 16 with $\lambda^* \in \arg\max_{\lambda\in\triangle_\mathcal{V}}\log\left(\det\left(\sum_{x\in\mathcal{V}}\lambda_x xx^\top + \gamma/TI\right)\right)$. So to summarize

$$\gamma_T := \sup_{\lambda\in\triangle_\mathcal{X}}\log\left(\det\left(T\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top + \gamma I\right)\right)$$

$$
\begin{aligned}
&\geq \sup_{\mathcal{V}\subset\mathcal{X}} \sup_{\lambda\in\triangle_{\mathcal{V}}} \log\left(\det\left(T\sum_{x\in\mathcal{V}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)\right)\\
&= \sup_{\mathcal{V}\subset\mathcal{X}} \log\left(\frac{\det\left(T\sum_{x\in\mathcal{V}}\lambda^*_x\phi(x)\phi(x)^\top+\gamma I\right)}{\det(\gamma I)}\right)+\log(\det(\gamma I))\\
&\geq \sup_{\mathcal{V}\subset\mathcal{X}} \frac{2}{3}\sup_{x\in\mathcal{V}}\|x\|^2_{(A(\lambda^*)+\gamma/TI)^{-1}}+\log(\det(\gamma I))\\
&\geq \frac{2}{3}\sup_{\mathcal{V}\subset\mathcal{X}}\inf_{\lambda\in\triangle_{\mathcal{V}}}\sup_{x\in\mathcal{X}}\|x\|^2_{(A(\lambda)+\gamma/TI)^{-1}}+d\log(\gamma)
\end{aligned}
$$

□

Corollary 4 (Consequence of Theorem 1 of (Degenne et al., 2020))**.** *Let τ_δ be the expected number of sample needed to find the best arm with probability at least $1-\delta$. For any $\theta_*\in\mathcal{E}$ we have*

$$
\liminf_{\delta\to 0}\frac{\mathbb{E}_{\theta_*}[\tau_\delta]}{\log(1/\delta)}\geq T^*(\theta_*)
$$

where

$$
T^{*-1}(\theta_*)=\max_{\lambda\in\Delta_{\mathcal{X}}}\inf_{x'\neq x_*}\sup_{\gamma\geq 0}F(\lambda,x',\gamma,\theta_*)
$$

$$
F(\lambda,x',\gamma,\theta_*)=\frac{\max\{(x'-x_*)^\top(A(\lambda)+\gamma I)^{-1}A(\lambda)\theta_*,0\}^2}{2\|x'-x_*\|^2_{(A(\lambda)+\gamma I)^{-1}}}+\frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)}-R^2\right)
$$

Proof. Recall theorem 1 of (Degenne et al., 2020). For any $\theta_*\in\mathcal{E}$ we have

$$
\liminf_{\delta\to 0}\frac{\mathbb{E}_{\theta_*}[\tau_\delta]}{\log(1/\delta)}\geq T^*(\theta_*)
$$

where we define the characteristic time through

$$
T^{*-1}(\theta_*):=\max_{\lambda\in\Delta_{\mathcal{X}}}\inf_{\theta\in\overline{S}_{x_*}}\|\theta-\theta_*\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x xx^\top)}
$$

with $\overline{S}_{x_*}=\{\theta\in\mathcal{E}\text{ s.t. }\exists x'\neq x_*,\theta^\top(x'-x_*)>0\}$ and with here $\mathcal{E}=\{\theta\in\mathbb{R}^d:\|\theta\|_2^2\leq R^2\}$.

We can now start the proof by writing $T^{*-1}(\theta)$ as

$$
T^{*-1}(\theta)=\max_{\lambda\in\Delta_{\mathcal{X}}}\inf_{x'\neq x_*}\inf_{\theta\in\mathcal{E},\theta^\top(x'-x_*)>0}\|\theta-\theta_*\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x xx^\top)}\,.
$$

Then, instead of

$$
\inf_{\theta\in\mathcal{E},\theta^\top(x'-x_*)>0}\|\theta-\theta_*\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x xx^\top)}
$$

we use $y=x'-x_*$ to write

$$
\inf_{\theta\in\mathbb{R}^d,\theta^\top y\geq 0,\|\theta\|_2^2\leq R^2}\frac{1}{2}\|\theta-\theta_*\|^2_{A(\lambda)}\,.
$$

We introduce the Lagrangian of this convex program

$$L(\theta, \gamma, \nu) = \frac{1}{2}\|\theta - \theta_*\|^2_{A(\lambda)} - \nu(\theta^\top y) + \frac{\gamma}{2}(\|\theta\|_2^2 - R^2) \ .$$

and solve

$$\inf_{\theta\in\mathbb{R}^d} L(\theta, \gamma, \nu) = \inf_{\theta\in\mathbb{R}^d} \frac{1}{2}\|\theta - \theta_*\|^2_{A(\lambda)} - \nu(\theta^\top y) + \frac{\gamma}{2}(\|\theta\|_2^2 - R^2)$$

$\theta \mapsto L(\theta, \gamma, \nu)$ is differentiable and convex so we compute the gradient

$$\nabla_\theta L(\theta, \gamma, \nu) = (A(\lambda) + \gamma I)\theta - A(\lambda)\theta_* - \nu y$$

and set it to zero to get

$$\widehat{\theta} = \arg\min_{\theta\in\mathbb{R}^d} L(\theta, \gamma, \nu) = (A(\lambda) + \gamma I)^{-1}(A(\lambda)\theta_* + \nu y) = \theta_* + (A(\lambda) + \gamma I)^{-1}(\nu y - \gamma\theta_*)$$

The cross term of both norms has absolute value $\gamma\nu\theta_*(A(\lambda) + \gamma I)^{-1}A(\lambda)y$, and they cancel. So we get

$$\begin{aligned}
L(\widehat{\theta}, \gamma, \nu) &= \frac{1}{2}\|(A(\lambda) + \gamma I)^{-1}(\nu y - \gamma\theta_*)\|^2_{A(\lambda)} - \nu y^\top \left((A(\lambda) + \gamma I)^{-1}(A(\lambda)\theta_* + \nu y)\right) \\
&\quad + \frac{\gamma}{2}(\|(A(\lambda) + \gamma I)^{-1}(A(\lambda)\theta_* + \nu y)\|_2^2 - R^2) \\
&= \frac{\gamma^2}{2}\|(A(\lambda) + \gamma I)^{-1}\theta_*\|^2_{A(\lambda)} + \frac{\gamma}{2}\|(A(\lambda) + \gamma I)^{-1}A(\lambda)\theta_*\|^2 - \frac{\gamma}{2}R^2 - \nu y^\top(A(\lambda) + \gamma I)^{-1}A(\lambda)\theta_* \\
&\quad + \frac{\nu^2}{2}\|(A(\lambda) + \gamma I)^{-1}y\|^2_{A(\lambda)} - \nu^2 y^\top(A(\lambda) + \gamma I)^{-1}y + \frac{\nu^2\gamma}{2}\|(A(\lambda) + \gamma I)^{-1}y\|^2 \\
&= \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right) - \nu y^\top(A(\lambda) + \gamma I)^{-1}A(\lambda)\theta_* - \frac{\nu^2}{2}\|y\|^2_{(A(\lambda)+\gamma I)^{-1}}
\end{aligned}$$

so

$$\sup_{\nu\geq 0} L(\widehat{\theta}, \gamma, \nu) = \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right) + \frac{(\max\{y^\top(A(\lambda) + \gamma I)^{-1}A(\lambda)\theta_*, 0\})^2}{2\|y\|^2_{(A(\lambda)+\gamma I)^{-1}}}$$

Conclusion:

$$\inf_{\theta\in\mathbb{R}^d, \theta^\top y\geq 0, \|\theta\|_2^2\leq R^2} \frac{1}{2}\|\theta - \theta_*\|^2_{A(\lambda)} = \sup_{\gamma\geq 0}\left\{\frac{(\max\{y^\top(A(\lambda) + \gamma I)^{-1}A(\lambda)\theta_*, 0\})^2}{2\|y\|^2_{(A(\lambda)+\gamma I)^{-1}}} + \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right)\right\}$$

Then for any $\theta_* \in \mathcal{E}$ we have

$$\begin{aligned}
\liminf_{\delta\to 0} \frac{\mathbb{E}_{\theta_*}[\tau_\delta]}{\log(1/\delta)} &\geq T^*(\theta_*) \\
&= \frac{1}{\max_{\lambda\in\Delta_\mathcal{X}} \inf_{x'\neq x_*} \sup_{\gamma\geq 0}\left\{\frac{(\max\{(x'-x_*)^\top(A(\lambda)+\gamma I)^{-1}A(\lambda)\theta_*, 0\})^2}{2\|x'-x_*\|^2_{(A(\lambda)+\gamma I)^{-1}}} + \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right)\right\}} \\
&= \inf_{\lambda\in\Delta_\mathcal{X}} \sup_{x'\neq x_*} \inf_{\gamma\geq 0} \frac{1}{F(\lambda, x', \gamma, \theta_*)}
\end{aligned}$$

With

$$F(\lambda, x', \gamma, \theta_*) := \frac{\max\{(x' - x_*)^\top (A(\lambda) + \gamma I)^{-1} A(\lambda)\theta_*, 0\}^2}{2\|x' - x_*\|^2_{(A(\lambda)+\gamma I)^{-1}}} + \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right)$$

□

Note that this result and its proof can be written in the case where ϕ is any feature map without any changes.

A.8 Experiments details

We briefly provide some additional details on the experiments. We used Python 3 and parallelized the simulations on a 2.9 GHz Intel Core i7. We computed the designs in each of the three experiments using mirror descent. We repeated the G-optimal design experiment 16 times, the kernels experiment 40 times, and the IPS vs. RIPS experiment 16 times. The G-optimal design experiment and the IPS vs. RIPS experiment used noise $\eta \sim N(0, 1)$ while in the kernels experiment used noise $\eta \sim N(0, 0.05)$. The confidence bounds in our plots are based on standard errors.

Appendix B

Appendix for Chap. 3

B.1 Summary of Gaussian Processes Approaches for Level Set Estimation

In Table B.2, we briefly summarize past algorithmic approaches to level set estimation. In general, past methods center around the design of an *acquisition function* which at each time t tells the algorithm which point to go sample. By contrast, the algorithms in this paper both use experimental design to to select batches of samples to go gather at one time.

Algorithm	**Acquisition Function**	**Theoretical guarantee**
Straddle	$\arg\max_i u_i(t) - \tau \wedge \tau - \ell_i(t)$	None, $u_i(t)$ and $\ell_i(t)$ are set as $1.96 \cdot \sigma_{t-1}$.
LSE	$\arg\max_i u_i(t) - \tau \wedge \tau - \ell_i(t)$	η-approximate solution in $T \lesssim \frac{\gamma_t \log(n/\delta)}{\eta^2}$
TruVar	$\arg\min_{x_i} \sum_{x_j} \sigma^2_{t-1\|x_i}(x_j)$	η-approximate solution in $T \lesssim \frac{\Gamma_t \log(n/\delta)}{\eta^2}$
RMILE	$\arg\max_{x_i}\{\mathbb{E}\sum(\mathbb{P}_{GP\|x_i}(f(x_j) > \tau) - \mathbb{P}_{GP}(f(x_j) > \tau)), \sigma^2(x_i)\}$	similar to A-optimality, no complexity guarantee
MELK	G-optimal design	Matching upper and lower bounds in the linear case.

Table B.1: Algorithms and theoretical guarantees for explicit LSE

B.2 Robust estimators for function means

In order for the algorithm to declare whether points x belong in G_α (or G_ϵ in the sequel) or not, we require an estimator of the function values $f(x)$. As we have introduced structure by assuming that f is well approximated by a function θ_* in the RKHS $\mathcal{H}$, we seek an estimator that leverages this structure to provide accurate estimates of many arms given samples of only a few. As a warmup, in the linear case where $\phi(\cdot)$ is the identity map, one could form the least squares or regularized least squares estimate of θ_* denoted $\widehat{\theta}$ and estimate the mean of any point x as $\widehat{\theta}^T x$. To sample to estimate θ_*, optimal design procedures first

Algorithm	**Acquisition function**	**Theoretical guarantee**
LSE-imp	$\arg\max_i \sigma^2(x_i)$	η-approximate solution in $T \lesssim \frac{\Gamma_t \log(n/\delta)}{\eta^2}$
MILK	XY optimal design over $\phi(x) - (1-\epsilon)\phi(x')$	Nearly matching upper and lower bounds.

Table B.2: Acquisition functions and theoretical guarantees for implicit level set estimation

compute a design $\lambda \in \triangle_{\mathcal{X}}$. Then for a specified number of samples N, it is common to use an efficient rounding procedure such as (Allen-Zhu et al., 2017) to compute an allocation of the N samples to the arms $\mathcal{X}$ such that x_i gets roughly $\lambda_i \cdot N$ samples (Fiez et al., 2019; Jun et al., 2020). Efficient rounding procedures require that $N = \Omega(d)$, and while this is a minor assumption in the case of a linear RKHS where $\phi(x) = x$, in general $\phi(x)$ may be infinite dimensional, and naive rounding is not possible. Instead of performing rounding given design λ, one may instead sample from λ directly and use inverse propensity scoring (IPS) which avoids bad dimensional factors but can have high variance.

In this work, we leverage the RIPS estimator from (Camilleri et al., 2021a) which combines IPS with robust mean estimation and regularization to control variance and is presented in Algorithm B.1. RIPS requires a robust mean estimator for its performance and theoretical guarantees. In Theorem 6, we state the guarantee of this estimator.

Algorithm B.1. `RIPS`: **R**obust **IPS** estimator

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $V \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ, regularization $\gamma > 0$, robust mean estimator $\widehat{\mu} : \mathbb{R}^* \to \mathbb{R}$

$$\lambda^* := \arg \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{v \in V} \|v\|_{\left(A^{(\gamma)}(\lambda)\right)^{-1}}$$

1: Randomly draw $\widetilde{x}_1, \ldots, \widetilde{x}_\tau$ from $\mathcal{X}$ according to λ^*
2: Set $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda^*)^{-1}\phi(\widetilde{x}_t)\widetilde{y}_t\}_{t=1}^{\tau})$
return $\widehat{\theta} := \arg\min_\theta \max_{v \in V} \frac{|\langle \theta, v\rangle - W^{(v)}|}{\|v\|_{A^{(\gamma)}(\lambda)^{-1}}}$

We next state the complete theoretical guarantee of the RIPS estimator.

Theorem 20 (Theorem 1, (Camilleri et al., 2021a))**.** *Consider the model $y = \langle \phi(x), \theta^* \rangle_{\mathcal{H}} + \zeta_x + \eta$ for misspecification $|\zeta_x| \leq h$ where it is assumed that $|y| \leq B$, $\mathbb{E}[\eta] = 0$, and $\mathbb{E}[\eta^2] \leq \sigma^2$. Fix any finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $V \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ and regularization $\gamma > 0$. If the RIPS procedure of Algorithm B.1 is run with $\frac{\delta}{|V|}$-robust mean estimator $\widehat{\mu}(\cdot)$ and if $\tau \geq c_1 \log(|V|/\delta)$ then with probability at least $1 - \delta$, we have*

$$\max_{v \in V} \frac{|W^{(v)} - \langle \theta_*, v\rangle|}{\|v\|_{\left(A^{(\gamma)}(\lambda)\right)^{-1}}} \leq \sqrt{\gamma}\|\theta_*\| + h + c\sqrt{\tfrac{(B^2+\sigma^2)}{\tau} \log(2|V|/\delta)}$$

Moreover, $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^{\tau})$ can be replaced by $\langle \widehat{\theta}, v\rangle$ by multiplying the RHS by a factor of 2.

For RIPS, we leverage Catoni's estimator (Lugosi and Mendelson, 2019) for which $c_1 = 2$ and $c = 4$ suffice.

B.3 Proofs for Explicit Level Set Estimation

B.3.1 Lower Bound

Proof of Theorem 5. Recall that we have assumed that $h = 0$ and $\phi(x) = x$. We begin with a result of (Fiez et al., 2019) that will be useful here.

Lemma 18 ((Fiez et al., 2019), Remark 2)**.** *The projection onto the closure of the set* $\{\theta \in \mathbb{R}^d : \theta^T x < \alpha\}$ *under the* $\|\cdot\|_{A(\lambda)}$ *norm is given by*

$$\theta_x := \theta - \frac{(\theta^T x - \alpha)A(\lambda)^{-1}x}{\|x\|^2_{A(\lambda)^{-1}}}.$$

By (Kaufmann et al., 2016), we have that the any δ-PAC algorithm for all-α requires

$$\min_{\lambda} \frac{KL(1-\delta,\delta)}{\min_{\theta' \in \text{Alt}(\theta_*)} \|\theta' - \theta_*\|_{A(\lambda)}}$$

where $\text{Alt}(\theta_*)$ is the set of alternates such that $G_\alpha(\theta_*) \neq G_\alpha(\theta')$ for any $\theta' \in \text{Alt}(\theta_*)$. The set of alternates may be decomposed as

$$\text{Alt}(\theta_*) = \left(\bigcup_{x \in G_\alpha(\theta_*)} \{\theta' : x \notin G_\alpha(\theta')\} \right) \cup \left(\bigcup_{x \in G_\alpha(\theta_*)^c} \{\theta' : x \in G_\alpha(\theta')\} \right)$$

Note that $x \in G_\alpha(\theta_*) \iff \theta_*^T x > \alpha$. Hence, the set of alternates for any $x \in G_\alpha(\theta_*)$ such that $x \in G_\alpha^c(\theta')$ for any $\theta' \in \text{Alt}(\theta_*)$ is given by

$$A_x := \{\theta \in \mathbb{R}^d : \theta^T x < \alpha\}.$$

Next note that $x \in G_\alpha^c(\theta_*) \iff \theta_*^T x < \alpha$. Hence, for any $x \in G_\alpha^c(\theta_*)$ the set of alternates such that $x \in G_\alpha(\theta')$ for any $\theta' \in \text{Alt}(\theta_*)$ is given by

$$A_x := \{\theta \in \mathbb{R}^d : \theta^T x > \alpha\}.$$

Next, we discuss how to project onto A_x. As this set is open, to be precise, we should take a point in the interior and consider the limit for a sequence approaching the boundary. For brevity, we simply project onto the closure and consider the closures of the A_x sets. Using the decomposition of $\text{Alt}(\theta_*)$ we have that

$$\min_{\theta' \in \text{Alt}(\theta_*)} \|\theta' - \theta_*\|_{A(\lambda)} = \min_x \min_{\theta' \in A_x} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\mathcal{S} \in \{G_\alpha, G_\alpha^c\}} \min_{x \in \mathcal{S}} \min_{\theta \in A_x} \|\theta' - \theta_*\|_{A(\lambda)}.$$

For $x \in G_\alpha(\theta_*)$, using Lemma 18 and recalling the definition of the set θ_x therein,

$$\min_{\theta' \in A_x} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\theta' \in \{\theta \in \mathbb{R}^d : \theta^T x \leq \alpha\}} \|\theta' - \theta_*\|_{A(\lambda)} = \|\theta_x - \theta_*\|_{A(\lambda)}.$$

The statement for points in G_α^c follows identically. Hence,

$$\min_{\theta' \in \text{Alt}(\theta_*)} \|\theta' - \theta_*\|_{A(\lambda)} = \min_x \|\theta_x - \theta_*\|_{A(\lambda)}$$

Note that

$$\|\theta_x - \theta_*\|_{A(\lambda)} = \frac{(\theta_*^T(x' - x) - \alpha)^2}{2\|x\|^2_{A(\lambda)^{-1}}}$$

by Theorem 2 of (Fiez et al., 2019). Hence, any δ-PAC algorithm requires at least

$$2 \min_{\lambda} \max_x \frac{\|x\|^2_{A(\lambda)^{-1}}}{(\theta_*^T x - \alpha)^2} KL(1-\delta,\delta)$$

samples in expectation. Noting that the binary entropy $KL(1-\delta,\delta) \geq \log(1/2.4\delta)$ completes the proof. $\square$

B.3.2 Upper Bound

Next, we restate Theorem 7 that bounds the complexity of MELK.

Theorem 21. *Fix $\delta > 0$, threshold $\alpha > 0$, tolerance $\widetilde{\beta}$, and regularization $\gamma \geq 0$. Define $\Delta_{\min}(\alpha) := \min |\phi(x)^T\theta_* - \alpha|$. Define also*

$$\bar{\beta}(\alpha) = \min\{\beta > 0 : 4(\sqrt{\gamma}\|\theta_*\| + h)(2 + \sqrt{f(\mathcal{X}, \{\phi(x)|x \in \mathcal{X}, |\phi(x)^T\theta_* - \alpha| \leq \beta\}; \gamma)}) \leq \beta\}.$$

With probability at least $1-\delta$, MELK returns a set $\widehat{\mathcal{R}} = (\mathcal{X} \setminus \widehat{B}_t)$ at time T_δ such that

$$\{x \in \mathcal{X} : f(x) \geq \alpha + \bar{\beta}(\alpha)\} \subseteq \widehat{\mathcal{R}} \subseteq \{x \in \mathcal{X} : f(x) \geq \alpha - \widetilde{\beta} - \bar{\beta}(\alpha)\}$$

and for any α, $\widetilde{\beta}$ such that $\max(\Delta_{\min}(\alpha), \widetilde{\beta}) \geq \bar{\beta}(\alpha)$

$$T_\delta \leq 256(B^2+\sigma^2) \min_{\lambda \in \tilde{\mathcal{X}}} \max_{x \in \mathcal{X}} \frac{\|\phi(x)\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{(\phi(x)^T\theta_* - \alpha)^2, \tilde{\beta}^2\}} \log\left(\frac{4|\mathcal{X}|^2 \lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil^2}{\delta}\right) + 2\log(|\mathcal{X}|/\delta)\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil$$

Recall the definition of the set $G_\alpha := \{x \in \mathcal{X} : f(x) > \alpha\}$.

Lemma 19. *For any $V \subset \mathcal{X}$ define $f(\mathcal{X}, V; \gamma) = \min_{\lambda \in \triangle_\mathcal{X}} \max_{v \in V} \|v\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}}$. In each round t, define the event*

$$E_t^c = \{|x^T(\widehat{\theta}_t - \theta_*)| \leq 2^{-t} + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)}\ \forall\ x \in \mathcal{A}_t\}$$

Holds $\mathbb{P}(\bigcup_{t=1}^\infty (E_t^c)^c) \leq \delta$.

Proof. Using Theorem 6, for any $x \in \mathcal{A}_t$ we have that with probability at least $1 - \delta_t/|\mathcal{X}|^2$

$$\begin{aligned}
|x^T(\widehat{\theta}_t - \theta_*)| &\leq \|x\|_{\left(\sum_{x \in \mathcal{X}} \lambda_x x x^T + \gamma I\right)^{-1}} \left(\sqrt{\gamma}\|\theta_*\| + h + c\sqrt{\tfrac{(B^2+\sigma^2)}{N_t}\log(2t^2|\mathcal{X}|^2/\delta)}\right) \\
&\leq \sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)}\left(\sqrt{\gamma}\|\theta_*\| + h + 2^{-t}/\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)}\right) \\
&\leq 2^{-t} + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)}
\end{aligned}$$

Since $|\mathcal{A}_t| \leq |\mathcal{X}|^2$, E_t^c holds for all $x \in \mathcal{A}_t$ with probability $1 - \delta_t$ via a union bound. Taking a second union bound over rounds, we have that

$$\mathbb{P}\left(\bigcup_{t=1}^\infty (E_t^c)^c\right) \leq \sum_{t=1}^\infty \mathbb{P}((E_t^c)^c) \leq \sum_{t=1}^\infty \delta_t = \sum_{t=1}^\infty \frac{\delta}{2t^2} \leq \delta$$

□

Define

$$\bar{t} = \max\{t : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{x \in \mathcal{X} : |x^T\theta_* - \alpha| \leq 2^{-t+2}\}; \gamma)}) \leq 2^{-t}\}.$$

As we will see in Lemmas 22 and 23,

$$\mathcal{A}_t \subset \left\{x \in \mathcal{X} : |x^T\theta_* - \alpha| \leq 2^{-t+1}\right\}.$$

Thus for $t \leq \bar{t}$, holds on $\bigcap_t E_t^c$ that

$$\forall x \in \mathcal{A}_t \ , \ |x^T(\widehat{\theta}_t - \theta_*)| \leq 2 \cdot 2^{-t}.$$

Lemma 20. *On $\bigcap_t E_t^c$, when $t \leq \bar{t}$ holds $\widehat{G}_t \subset G_\alpha^\phi := \{x \ : \ \phi(x)^T\theta_* > \alpha\}$.*

Remark: If $h = 0$, $G_\alpha^\phi = G_\alpha$.

Proof.

$$\begin{aligned}
x \in \widehat{G}_t &\iff \exists t' \leq t \ : \ \phi(x)^T\widehat{\theta}_{t'} \geq \alpha + 2 \cdot 2^{-t'} \\
&\iff \exists t' \leq t \ : \ \phi(x)^T(\widehat{\theta}_{t'} - \theta_*) + \phi(x)^T\theta_* \geq \alpha + 2 \cdot 2^{-t'} \\
&\overset{\bigcap_t E_t^c}{\Longrightarrow} \phi(x)^T\theta_* > \alpha \\
&\iff x \in G_\alpha^\phi.
\end{aligned}$$

□

Lemma 21. *On $\bigcap_t E_t^c$, when $t \leq \bar{t}$ holds, $\widehat{B}_t \subset (G_\alpha^\phi)^c$.*

Proof.

$$\begin{aligned}
x \in \widehat{B}_t \iff &\exists t' \leq t \ : \ \phi(x)^T\widehat{\theta}_t \leq \alpha - 2 \cdot 2^{-t'} \\
&\exists t' \leq t \ : \ \phi(x)^T(\widehat{\theta}_t - \theta_*) + \phi(x)^T\theta_* \leq \alpha - 2 \cdot 2^{-t'} \\
&\overset{\bigcap_t E_t^c}{\Longrightarrow} \phi(x)^T\theta_* < \alpha \\
&\iff x \in (G_\alpha^\phi)^c.
\end{aligned}$$

□

Lemma 22. *On the event $\bigcap_t E_t^c$, when $t \leq \bar{t}$ holds,*

$$\mathcal{A}_t \cap G_\alpha^\phi \subset \left\{x \in G_\alpha^\phi \middle| |\phi(x)^T\theta_* - \alpha| \leq 2^{-t+2}\right\} =: \mathcal{S}_t^{\textit{Above}}$$

Proof. For any $x \in G_\alpha^\phi$ such that $\phi(x)^T\theta_* > \alpha + 2^{-t+1}$, if $t \geq \log(4(\alpha - \phi(x)^T\theta_*)^{-1})$ and $t \leq \bar{t}$, then

$$\phi(x)^T\widehat{\theta}_t = \phi(x)^T(\widehat{\theta}_t - \theta_*) + \phi(x)^T\theta_* > -2^{-t+1} + \alpha + 2^{-t+1} = \alpha \geq \alpha$$

which implies that $x \in \widehat{G}_t$. Noting that $\mathcal{A}_t \cap \widehat{G}_{t-1} = \emptyset$ completes the proof. □

Lemma 23. *On the event $\bigcap_t E_t^c$, when $t \leq \bar{t}$ holds,*

$$\mathcal{A}_t \cap (G_\alpha^\phi)^c \subset \left\{x \in (G_\alpha^\phi)^C \middle| |\phi(x)^T\theta_* - \alpha| \leq 2^{-t+2}\right\} =: \mathcal{S}_t^{\textit{Below}}$$

Proof. The proof follows identically as that of Lemma 22 □

Remark: Lemmas 22 and 23 jointly imply that $\mathcal{A}_t \subset \left\{x|\phi(x)^T\theta_* - \alpha| \leq 2^{-t+2}\right\} =: \mathcal{S}_t$ for $t \leq \bar{t}$. Furthermore, $f(\mathcal{X}, \mathcal{A}_t, \gamma) \leq f(\mathcal{X}, \mathcal{S}_t, \gamma)$.

Remark:
The algorithm stops on either of two conditions. On one hand if $t \geq \lceil \log_2(4/\widetilde{\beta}) \rceil =: t_\beta$, then it has achieved precision $\widetilde{\beta}$ as desired and it terminates. Otherwise, it terminates if $\widehat{G}_t \cup \widehat{B}_t = \mathcal{X}$. This occurs when $\widetilde{\beta}$ is very small. Define $\Delta_{\min}(\alpha) := \min |\phi(x)^T\theta_* - \alpha|$. Recall

$$\begin{aligned}
\bar{t} &= \max\{t : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{x \in \mathcal{X} : |\phi(x)^T\theta_* - \alpha| \leq 4 \cdot 2^{-t}\}; \gamma)}) \leq 2^{-t}\} \\
&= \max\{t : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{x \in \mathcal{X} : |\phi(x)^T\theta_* - \alpha| \leq 4 \cdot 2^{-t}\}; \gamma)}) \leq 4 \cdot 2^{-t}\} \\
&= -2 + \max\{t : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{x \in \mathcal{X} : |\phi(x)^T\theta_* - \alpha| \leq 2^{-t}\}; \gamma)}) \leq 2^{-t}\} \\
&= -3 - \log_2(\min\{\beta > 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{x \in \mathcal{X} : |\phi(x)^T\theta_* - \alpha| \leq \beta\}; \gamma)}) \leq \beta\}).
\end{aligned}$$

This defines

$$\bar{\beta} = \min\{\beta > 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{x \in \mathcal{X} : |\phi(x)^T\theta_* - \alpha| \leq \beta\}; \gamma)}) \leq \beta\}.$$

Let $t_{\max}$ denote the random variable of the last round before the algorithm terminates. The following Lemmas give a guarantee on the set $\mathcal{X} \setminus \widehat{B}_t$ at termination.

Lemma 24. *On the event $\bigcap_{t=1}^{\infty} E_t^c$, MELK returns a set $(\mathcal{X} \setminus \widehat{B}_{t_{\max}})$ such that $\{x : f(x) > \alpha + \bar{\beta}(\alpha)\} \subset (\mathcal{X} \setminus \widehat{B}_{t_{\max}})$.*

Proof. Take any x such that $f(x) > \alpha + \bar{\beta}(\alpha)$ and recall that by assumption $|f(x) - \phi(x)^T\theta_*| \leq h$ for all $x \in \mathcal{X}$. We consider two cases. In the first case, assume that $t_{\max} \leq \bar{t}$. We claim that in this case $\nexists t$ such that $x \in \widehat{B}_t$. We prove this by contradiction. Assume not. Then $\exists t$ such that

$$\begin{aligned}
\widehat{\theta}_t^T\phi(x) < \alpha - 2^{-t+1} &\iff \phi(x)^T(\widehat{\theta}_t - \theta_*) + \phi(x)^T\theta_* < \alpha - 2^{-t+1} \\
&\overset{E_t^c}{\implies} -2^{-t} - (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)} + \phi(x)^T\theta_* < \alpha - 2^{-t+1} \\
&\implies -2^{-t} - (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)} + \phi(x)^T\theta_* < \alpha - 2^{-t+1} \\
&\implies -(\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)} + f(x) - h < \alpha - 2^{-t} \\
&\implies f(x) < \alpha - 2^{-t} + h + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}.
\end{aligned}$$

Recalling that we have assumed that $f(x) > \alpha + \bar{\beta}(\alpha)$. Hence, this implies that

$$\bar{\beta}(\alpha) < -2^{-t} + h + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}.$$

Note that $\bar{\beta}(\alpha) > 0$. As we have assumed that, $t \leq t_{\max} \leq \bar{t}$, we have that $2^{-t} \geq \left(\sqrt{\gamma}\|\theta_*\| + h\right)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}$ using the definition of $\bar{t}$. Hence, we have that

$$h > \bar{\beta}(\alpha) > 4h$$

which is a contradiction where the final inequality follows from the definition of $\bar{\beta}(\alpha)$ for $\gamma > 0$. Hence, in this case we have shown that $\{x : f(x) > \alpha + \bar{\beta}(\alpha)\} \subset (\mathcal{X} \setminus \widehat{B}_{t_{\max}})$.

In the second case, assume that $t_{\max} > \bar{t}$ and take x such that $f(x) > \alpha + \bar{\beta}(\alpha)$. We claim that $x \in \widehat{G}_{\bar{t}}$ and hence $x \notin \mathcal{A}_t$ for any $t > \bar{t}$ and thus is never added to $\widehat{B}_t$. This occurs if

$$\begin{aligned}
\phi(x)^T\widehat{\theta}_{\bar{t}} > \alpha + 2^{-\bar{t}+1} &\iff \phi(x)^T(\widehat{\theta}_{\bar{t}} - \theta_*) + \phi(x)^T\theta_* > \alpha + 2^{-\bar{t}+1} \\
&\overset{E_{\bar{t}}^c}{\impliedby} -2^{-\bar{t}} - (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_{\bar{t}}; \gamma)} + \phi(x)^T\theta_* \geq \alpha + 2^{-\bar{t}+1} \\
&\impliedby -2^{-\bar{t}} - (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_{\bar{t}}; \gamma)} + \phi(x)^T\theta_* \geq \alpha + 2^{-\bar{t}+1} \\
&\iff \phi(x)^T\theta_* \geq \alpha + 2^{-\bar{t}+1} + 2^{-\bar{t}} + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_{\bar{t}}; \gamma)}
\end{aligned}$$

Recall that $f(x) > \alpha + \bar{\beta}(\alpha)$. Furthermore, we have by the definition of $\bar{t}$ that

$$2^{-\bar{t}} \geq (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_{\bar{t}}; \gamma)}.$$

Hence, the above is implied by $\bar{\beta}(\alpha) - h \geq 4 \cdot 2^{-t} = 0.5\bar{\beta}(\alpha)$ where the final equality holds by definition of $\bar{\beta}(\alpha)$. Noting that $\bar{\beta}(\alpha) > 4h$ proves this claim. In summary, we have shown that for any x such that $f(x) > \alpha + \bar{\beta}(\alpha)$, if $t_{\max} \leq \bar{t}$, then x is never added to $\widehat{B}_t$ and hence is contained in the set $\mathcal{X} \setminus \widehat{B}_t$ at termination, and if otherwise that $t_{\max} > \bar{t}$, then x is added to the set $\widehat{G}_t$ before round $\bar{t} + 1$ and hence is removed from the active set and never added to $\widehat{B}_t$. Applying this argument to any x such that $f(x) > \alpha + \bar{\beta}(\alpha)$ completes the proof. □

Lemma 25. *On the event $\bigcap_{t=1}^{\infty} E_t^c$, MELK returns a set $(\mathcal{X} \setminus \widehat{B}_{t_{\max}})$ such that $(\mathcal{X} \setminus \widehat{B}_{t_{\max}}) \subset \{x : f(x) > \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}\}$.*

Proof. Take any x such that $f(x) < \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}$. We claim that there exists a $t \leq t_{\max}$ such that x is added to $\widehat{B}_t$ which implies that $x \notin (\mathcal{X} \setminus \widehat{B}_{t_{\max}})$. Suppose for contradiction that this is not the case. Then for all $t \leq t_{\max}$,

$$\begin{aligned}
\widehat{\theta}_t^T\phi(x) > \alpha - 2^{-t+1} &\iff \phi(x)^T(\widehat{\theta}_t - \theta_*) + \phi(x)^T\theta_* > \alpha - 2^{-t+1} \\
&\overset{E_t^c}{\implies} 2^{-t} + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)} + \phi(x)^T\theta_* > \alpha - 2^{-t+1} \\
&\implies 2^{-t} + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)} + \phi(x)^T\theta_* > \alpha - 2^{-t+1} \\
&\implies (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)} + f(x) + h > \alpha - 2^{-t+1} - 2^{-t} \\
&\implies f(x) > \alpha - 2^{-t+1} - 2^{-t} - h - (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}.
\end{aligned}$$

Plugging in $f(x) < \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}$, the above implies

$$\bar{\beta}(\alpha) + \widetilde{\beta} < 2^{-t+1} + 2^{-t} + h + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)} \tag{B.1}$$

Next, recall that MELK terminates either on the condition that $t = \lceil \log_2(4/\widetilde{\beta}) \rceil$ or that $\widehat{G}_t \cup \widehat{B}_t = \mathcal{X}$. Using this, we brake our analysis into cases.

Case 1: $t_{\max} = \lceil \log_2(4/\widetilde{\beta}) \rceil \leq \bar{t}$.

In this case, MELK stops due to the $\widetilde{\beta}$ tolerance in a round before $\bar{t}$. For $t \leq \bar{t}$, we have that $2^{-t} \geq + \left(\sqrt{\gamma}\|\theta_*\| + h\right)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}$. Hence, the above implies that

$$\bar{\beta}(\alpha) + \widetilde{\beta} < 2^{-t+2} + h.$$

As we have assumed this condition for all $t \leq t_{\max}$, we may plug in $t_{\max}$ which implies

$$\bar{\beta}(\alpha) + \widetilde{\beta} < \widetilde{\beta} + h.$$

As $\bar{\beta}(\alpha) > h$, this is a contradiction. Hence there must exist a t such that $x \in \widehat{B}_t$.

Case 2: $t_{\max} \leq \bar{t} < \lceil \log_2(4/\widetilde{\beta}) \rceil$.

In this case, MELK terminates before round $t = \lceil \log_2(4/\widetilde{\beta}) \rceil$. Hence, it does so on the condition that $\widehat{G}_t \cup \widehat{B}_t = \mathcal{X}$. Note that for $f(x) < \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}$, we have that $x \in (G_\alpha^\phi)^c$ since $\bar{\beta}(\alpha) > h$ and $\widetilde{\beta} \geq 0$. If we terminate before round $\bar{t}$, we have by Lemma 21 that $(G_\alpha^\phi)^c \subset \widehat{B}_t$ which implies that $x \in \widehat{B}_{t_{\max}}$. This contradicts the assumption that $\nexists t : x \in \widehat{B}_t$.

Case 3: $\bar{t} < t_{\max}$.

In this case, MELK terminates at a round after $\bar{t}$. In this setting, we argue that $x \in \widehat{B}_{\bar{t}}$. Recall that for any $t \leq \bar{t}$, (B.1) simplifies to

$$\bar{\beta}(\alpha) + \widetilde{\beta} < 2^{-t+2} + h$$

Plugging in $\bar{t}$, and noting that $2^{-\bar{t}+2} = \frac{1}{2}\bar{\beta}(\alpha)$, the above implies

$$\bar{\beta}(\alpha) + \widetilde{\beta} < \frac{1}{2}\bar{\beta}(\alpha) + h.$$

Noting that $\bar{\beta}(\alpha) > 4h$, shows that the above is a contradiction. Hence, there exists a $t \leq \bar{t}$ such that $x \in \widehat{B}_t$.

Therefore, in all cases we have shown that for any x such that $f(x) < \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}$, $x \in \widehat{B}_t$. Therefore, for the returned set $\mathcal{X} \setminus \widehat{B}_{t_{\max}}$, we have that

$$(\mathcal{X} \setminus \widehat{B}_{t_{\max}}) \subset \{x : f(x) > \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}\}.$$

□

Proof of Theorem 7. Throughout, assume the high probability event $\bigcap_T E_t^c$. By Lemmas 24 and 25 in conjunction with the high probability event $\bigcap E_t^c$ we have correctness. It remains to control the sample complexity of MELK. Recall that we have assumed that $\max(\Delta_{\min}(\alpha), \widetilde{\beta}) \geq \bar{\beta}(\alpha)$. This implies that $\min\{\lceil \log_2(4/\Delta_{\min}(\alpha)) \rceil, \lceil \log_2(4/\widetilde{\beta}) \rceil\} \leq \bar{t}$. Applying Lemmas 22 and 23, we have that $t_{\max} \leq \min\{\lceil \log_2(4/\Delta_{\min}(\alpha)) \rceil, \lceil \log_2(4/\beta) \rceil\} \leq \bar{t}$ and that $\mathcal{A}_t \subseteq \mathcal{S}_t$ for all rounds t. Now we proceed by bounding the total number of samples drawn.

$$\begin{aligned}
\tau &\leq \sum_{t=1}^{t_{\max}} N_t \\
&\leq \sum_{t=1}^{\min\{t \geq \lceil \log_2(4/\Delta_{\min}(\alpha)) \rceil, t \geq \lceil \log_2(4/\widetilde{\beta}) \rceil\}} N_t \\
&= \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\alpha) Vee \widetilde{\beta})^{-1} \rceil} \max\left\{c_1 \log(|\mathcal{X}|/\delta), c^2 2^{2t} f(\mathcal{A}_t; \gamma)(B^2 + \sigma^2) \log(2t^2|\mathcal{X}|^2/\delta)\right\} \\
&\leq c_1 \log(|\mathcal{X}|/\delta) \lceil \log_2(4(\Delta_{\min}(\alpha) Vee \widetilde{\beta})^{-1} \rceil + c^2(B^2 + \sigma^2) \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\alpha) Vee \widetilde{\beta})^{-1} \rceil} 2^{2t} f(\mathcal{A}_t; \gamma) \cdot \log(2t^2|\mathcal{X}|^2/\delta) \\
&= c_1 \log(|\mathcal{X}|/\delta) \lceil \log_2(4(\Delta_{\min}(\alpha) Vee \widetilde{\beta})^{-1} \rceil + \\
&\qquad c^2(B^2 + \sigma^2) \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\alpha) Vee \widetilde{\beta})^{-1} \rceil} 2^{2t} \min_{\lambda \in \widetilde{X}} \max_{x \in \mathcal{A}_t} \|x\|^2_{\left(\sum_{x \in \mathcal{X}} \lambda_t(x) \phi(x) \phi(x)^T + \gamma I\right)^{-1}} \cdot \log(2t^2|\mathcal{X}|^2/\delta)
\end{aligned}$$

$$
\begin{aligned}
&\leq c_1 \log(|\mathcal{X}|/\delta)\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil + \\
&\qquad c^2(B^2+\sigma^2)\log\left(\frac{4|\mathcal{X}|^2\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil^2}{\delta}\right) \\
&\qquad \cdot \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil} 2^{2t} \min_{\lambda\in\widetilde{X}} \max_{x\in\mathcal{A}_t} \|x\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_t(x)\phi(x)\phi(x)^T+\gamma I\right)^{-1}} \\
&\overset{\mathcal{A}_t\subset\mathcal{S}_t}{\leq} c_1 \log(|\mathcal{X}|/\delta)\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil + \\
&\qquad c^2(B^2+\sigma^2)\log\left(\frac{4|\mathcal{X}|^2\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil^2}{\delta}\right) \\
&\qquad \cdot \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil} 2^{2t} \min_{\lambda\in\widetilde{X}} \max_{x\in\mathcal{S}_t} \|x\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_t(x)\phi(x)\phi(x)^T+\gamma I\right)^{-1}}.
\end{aligned}
$$

It remains to control the final summation. To do so, note that

$$
\begin{aligned}
&\frac{1}{\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil} \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil} 2^{2t} \min_{\lambda\in\widetilde{X}} \max_{x\in\mathcal{S}_t} \|x\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_t(x)\phi(x)\phi(x)^T+\gamma I\right)^{-1}} \\
&\leq \max_{t\leq\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil} \min_{\lambda\in\widetilde{X}} 2^{2t} \min_{\lambda\in\widetilde{X}} \max_{x\in\mathcal{S}_t} \|x\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_t(x)\phi(x)\phi(x)^T+\gamma I\right)^{-1}} \\
&\leq \min_{\lambda\in\widetilde{X}} \max_{t\leq\lceil \log_2(4(\Delta_{\min}(\alpha)Vee\widetilde{\beta})^{-1}\rceil} 2^{2t} \min_{\lambda\in\widetilde{X}} \max_{x\in\mathcal{S}_t} \|x\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_t(x)\phi(x)\phi(x)^T+\gamma I\right)^{-1}} \\
&\leq 16 \min_{\lambda\in\widetilde{X}} \max_{x} \frac{\|\phi(x)\|^2_{(\sum_{x\in\mathcal{X}}\lambda_t(x)\phi(x)\phi(x)^T+\gamma I)^{-1}}}{\max\{(\phi(x)^T\theta_* - \alpha)^2, \widetilde{\beta}^2\}}
\end{aligned}
$$

Plugging this along with $c = 4$ and $c_1 = 2$ for Theorem 20 from RIPS with the Catoni estimator in completes the proof. □

B.4 Proofs for Implicit Level Set Estimation

B.4.1 Lower Bounds

Proof of Theorem 8. Recall that in this setting, $h = 0$ and $\phi(\boldsymbol{x}) = \boldsymbol{x}$. By (Kaufmann et al., 2016), we have that the any δ-PAC algorithm for all-ϵ requires

$$
\min_{\lambda} \frac{KL(1-\delta,\delta)}{\min_{\theta'\in \text{Alt}(\theta_*)} \|\theta' - \theta_*\|_{A(\lambda)}}
$$

where $\text{Alt}(\theta_*)$ is the set of alternates such that $G_\epsilon(\theta_*) \neq G_\epsilon(\theta')$ for any $\theta' \in \text{Alt}(\theta_*)$. The set of alternates may be decomposed as

$$
\textbf{Alt}(\theta_*) = \left(\bigcup_{\boldsymbol{x}\in G_\epsilon(\theta_*)} \{\theta' : \boldsymbol{x} \notin G_\epsilon(\theta')\}\right) \cup \left(\bigcup_{\boldsymbol{x}\in G_\epsilon(\theta_*)^c} \{\theta' : \boldsymbol{x} \in G_\epsilon(\theta')\}\right)
$$

By Lemma 4, $\boldsymbol{x} \in G_\epsilon \iff \forall \boldsymbol{x}' : \theta_*^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}') > 0$. Hence, the set of alternates for any $\boldsymbol{x} \in G_\epsilon(\theta_*)$ such that $\boldsymbol{x} \in G_\epsilon^c(\theta')$ for any $\theta' \in \mathbf{Alt}(\theta_*)$ is given by

$$A_{\boldsymbol{x}} := \bigcup_{\boldsymbol{x}' \in \mathcal{X}} \{\theta \in \mathbb{R}^d : \theta^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}') < 0\}.$$

Furthermore, by Lemma 4 $\boldsymbol{x} \in G_\epsilon^c \iff \exists \boldsymbol{x}' : \theta_*^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}') < 0$. Hence, for any $\boldsymbol{x} \in G_\epsilon^c(\theta_*)$ the set of alternates such that $\boldsymbol{x} \in G_\epsilon(\theta')$ for any $\theta' \in \mathbf{Alt}(\theta_*)$ is given by

$$A_{\boldsymbol{x}} := \bigcap_{\boldsymbol{x}' \in \mathcal{X}} \{\theta \in \mathbb{R}^d : \theta^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}') > 0\}.$$

Next, we discuss how to project onto $A_{\boldsymbol{x}}$. As this set is open, to be precise, we should take a point in the interior and consider the limit for a sequence approaching the boundary. For brevity, we simply project onto the closure and consider the closures of the $A_{\boldsymbol{x}}$ sets. Using the decomposition of $\mathbf{Alt}(\theta_*)$ we have that

$$\min_{\theta' \in \text{Alt}(\theta_*)} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\boldsymbol{x}} \min_{\theta' \in A_{\boldsymbol{x}}} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\mathcal{S} \in \{G_\epsilon, G_\epsilon^c\}} \min_{\boldsymbol{x} \in \mathcal{S}} \min_{\theta \in A_{\boldsymbol{x}}} \|\theta' - \theta_*\|_{A(\lambda)}.$$

Reminiscent of Lemma 18, we define

$$\theta^\epsilon_{\boldsymbol{x},\boldsymbol{x}'}(\lambda) := \theta_* - [\theta_*^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}')] \frac{\mathcal{A}(\lambda)^{-1}(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'))}{\|\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'\|^2_{\mathcal{A}(\lambda)^{-1}}}.$$

For $\boldsymbol{x} \in G_\epsilon(\theta_*)$, using Lemma 18,

$$\min_{\theta' \in A_{\boldsymbol{x}}} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\theta' \in \bigcup_{\boldsymbol{x}' \in \mathcal{X}} \{\theta \in \mathbb{R}^d : \theta^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}') < 0\}} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\boldsymbol{x}'} \|\theta^\epsilon_{\boldsymbol{x},\boldsymbol{x}'}(\lambda) - \theta_*\|_{A(\lambda)}$$

where the latter equality follows since projecting onto a union of hyperplanes is achieved by the projection onto the closest constituent.

For $\boldsymbol{x} \in G_\epsilon^c(\theta_*)$ note that $A_{\boldsymbol{x}}$ is an intersection of half spaces $\{\theta \in \mathbb{R}^d : \theta^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}) > 0\}$ for $\boldsymbol{x}' \in \mathcal{X}$. As it is not in general possible to give a closed form expression for projection onto an intersection of convex sets. However, we may at a (possibly very loose) minimum note that the projection onto the union of the hyperplanes is at least as far as the projection onto the furthest hyperplane. Therefore, for any $\boldsymbol{x} \in G_\epsilon(\theta_*)^c$,

$$\min_{\theta' \in A_{\boldsymbol{x}}} \|\theta' - \theta_*\|_{A(\lambda)} = \min_{\theta' \in \bigcap_{\boldsymbol{x}' \in \mathcal{X}} \{\theta \in \mathbb{R}^d : \theta^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}') > 0\}} \|\theta' - \theta_*\|_{A(\lambda)} \leq \max_{\boldsymbol{x}'} \|\theta^\epsilon_{\boldsymbol{x},\boldsymbol{x}'}(\lambda) - \theta_*\|_{A(\lambda)}$$

Hence we have that

$$\min_{\theta' \in \text{Alt}(\theta_*)} \|\theta' - \theta_*\|_{A(\lambda)} \leq \min \left\{ \min_{\boldsymbol{x} \in G_\epsilon} \min_{\boldsymbol{x}'} \|\theta^\epsilon_{\boldsymbol{x},\boldsymbol{x}'}(\lambda) - \theta_*\|_{A(\lambda)}, \min_{\boldsymbol{x} \in G_\epsilon^c} \max_{\boldsymbol{x}'} \|\theta^\epsilon_{\boldsymbol{x},\boldsymbol{x}'}(\lambda) - \theta_*\|_{A(\lambda)} \right\}.$$

Note that

$$\|\theta^\epsilon_{\boldsymbol{x},\boldsymbol{x}'}(\lambda) - \theta_*\|_{A(\lambda)} = 2\frac{(\theta_*^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'))^2}{\|\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'\|^2_{A(\lambda)^{-1}}}$$

by Theorem 2 of (Fiez et al., 2019). Hence, any δ-PAC algorithm requires

$$2 \min_\lambda \max \left\{ \max_{\boldsymbol{x} \in G_\epsilon} \max_{\boldsymbol{x}'} \frac{\|\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'\|^2_{A(\lambda)^{-1}}}{(\theta_*^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'))^2}, \max_{\boldsymbol{x} \in G_\epsilon^c} \min_{\boldsymbol{x}'} \frac{\|\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'\|^2_{A(\lambda)^{-1}}}{(\theta_*^T(\boldsymbol{x} - (1-\epsilon)\boldsymbol{x}'))^2} \right\} KL(1-\delta, \delta)$$

samples in expectation. Noting that $KL(1-\delta, \delta) \geq \log(1/2.4\delta)$ completes the proof. □

B.4.2 Comparison to the lower bound of (Mason et al., 2020)

Here, we compare the sample complexity given in Theorem 9 to the result of Mason et al., (Mason et al., 2020) studying the problem of finding all ϵ-good arms in multi-armed bandits. Our setting captures this problem in the special case that $\phi(\boldsymbol{x}) = \boldsymbol{x}$, $\boldsymbol{x}_i = e_i \in \mathbb{R}^{|\mathcal{X}|}$, $h = 0$, and $\widetilde{\beta} = 0$. Additionally, take $\gamma \to 0$. For consistency with the notation of (Mason et al., 2020), let $\mu_i = f(\boldsymbol{x}_i)$ and $|\mathcal{X}| = n$. In this setting, the problem of implicit level set estimation reduces to identifying the set $\{i : \mu_i > (1-\epsilon)\mu_1\}$ where we assume without loss of generality that the means are sorted in descending order such that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$.

Lemma 26. *The term $H^{MILK}(\theta_*) = cH_{(ST)^2}$ for a constant c where $H_{(ST)^2}$ is the complexity parameter of the $(ST)^2$ algorithm from (Mason et al., 2020).*

In particular, (Mason et al., 2020) show in Theorem 4.1 that a complexity of $H_{(ST)^2}$ is optimal up to logarithmic factors for any fixed δ via a moderate confidence bound. This exceeds the lower bound given in Theorem 8 specialized to this case. In particular, this highlights that the lower bound given in Theorem 8 is not achievable except possibly as $\delta \to 0$. Instead, we show that MILK achieves the optimal non-asymptotic sample complexity for finding all ϵ-good arms.

Proof of Lemma 26. First, we recall some notation from (Mason et al., 2020) necessary for this lemma. Let $\tilde{\alpha}_\epsilon = \min_{i \in G_\epsilon} \mu_i - (1-\epsilon)\mu_1$ and let $\tilde{\beta}_\epsilon = \min_{i \in G_\epsilon^c}(1-\epsilon)\mu_1 - \mu_i$. For brevity, we let $k = \arg\min_{i \in G_\epsilon} \mu_i$ and $k+1 = \arg\max_{i \in G_\epsilon^c} \mu_i$ where we take $n > k$. If this condition does not hold the same argument as below suffices ignoring all terms in G_ϵ^c. Hence we have that $\frac{\mu_k}{1-\epsilon} = \mu_1 + \frac{\tilde{\alpha}_\epsilon}{1-\epsilon}$ and $\frac{\mu_{k+1}}{1-\epsilon} = \mu_1 - \frac{\tilde{\beta}_\epsilon}{1-\epsilon}$. Furthermore, (Mason et al., 2020) restrict to the case of $\epsilon \in [1/2, 1)$.

We begin by lower bounding the complexity parameter $H^{\texttt{MILK}}(\theta_*)$. We analyze the two terms given in Theorem 9, $H^{\texttt{MILK1}}$ and $H^{\texttt{MILK2}}$ separately. $H^{\texttt{MILK1}}$ reduces to

$$\begin{aligned}
\max_{e_i \in G_\epsilon} \max_{e_j} \frac{\|e_j - e_i\|^2_{A(\lambda)^{-1}}}{(\mu_i - (1-\epsilon)\mu_j)^2} &= \max_{e_i \in G_\epsilon} \max_{e_j} \frac{1/\lambda_i + 1/\lambda_j}{(\mu_i - (1-\epsilon)\mu_j)^2} \\
&\geq \max\left\{ \max_{e_i \in G_\epsilon} \frac{1/\lambda_i}{(\mu_i - (1-\epsilon)\mu_1)^2}, \max_{e_j} \frac{1/\lambda_j}{(\frac{\mu_k}{1-\epsilon} - \mu_j)^2} \right\} \\
&= \max\left\{ \max_{e_i \in G_\epsilon} \frac{1/\lambda_i}{(\mu_1 - \mu_i - \epsilon)^2}, \max_{e_j} \frac{1/\lambda_j}{(\mu_1 + \frac{\tilde{\alpha}_\epsilon}{1-\epsilon} - \mu_j)^2} \right\}
\end{aligned}$$

where the final step follows by the definition of $\tilde{\alpha}_\epsilon$. The penultimate step follows by first maximizing over $i \in G_\epsilon$ which introduces a factor of μ_k. Then we may multiply the denominator by $(1-\epsilon)^2/(1-\epsilon)^2$ and upper bound $(1-\epsilon)^2 \leq 0.25 < 1$ since $\epsilon \geq 1/2$ to achieve the result.

$H^{\texttt{MILK2}}$ reduces to

$$\begin{aligned}
\max_{e_i \in G_\epsilon^c} \max_{e_j} \frac{\|e_j - e_i\|^2_{A(\lambda)^{-1}}}{((1-\epsilon)\mu_1 - \mu_i)^2} &= \max_{e_i \in G_\epsilon^c} \max_{e_j} \frac{1/\lambda_i + 1/\lambda_j}{((1-\epsilon)\mu_1 - \mu_i)^2} \\
&\geq \max\left\{ \max_{e_i \in G_\epsilon^c} \frac{1/\lambda_i}{((1-\epsilon)\mu_1 - \mu_i)^2}, \max_{e_i \in G_\epsilon^c} \max_{e_j} \frac{1/\lambda_j}{((1-\epsilon)\mu_1 - \mu_i)^2} \right\} \\
&\geq \max\left\{ \max_{e_i \in G_\epsilon^c} \frac{1/\lambda_i}{((1-\epsilon)\mu_1 - \mu_i)^2}, \max_{e_j} \frac{1/\lambda_j}{((1-\epsilon)\mu_1 - \mu_{k+1})^2} \right\} \\
&\geq \max\left\{ \max_{e_i \in G_\epsilon^c} \frac{1/\lambda_i}{((1-\epsilon)\mu_1 - \mu_i)^2}, \max_{e_j} \frac{1/\lambda_j}{(\mu_1 - \frac{\mu_{k+1}}{1-\epsilon})^2} \right\}
\end{aligned}$$

$$= \max\left\{ \max_{e_i \in G_\epsilon^c} \frac{1/\lambda_i}{((1-\epsilon)\mu_1 - \mu_i)^2}, \max_{e_j} \frac{1/\lambda_j}{\left(\left(\mu_1 - \frac{\tilde{\beta}_\epsilon}{1-\epsilon}\right) - \mu_1\right)^2} \right\}$$

$$= \max\left\{ \max_{e_i \in G_\epsilon^c} \frac{1/\lambda_i}{((1-\epsilon)\mu_1 - \mu_i)^2}, \max_{e_j} \frac{1/\lambda_j}{\left(\left(\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon}\right) - \mu_1\right)^2} \right\}$$

$$\geq \max\left\{ \max_{e_i \in G_\epsilon^c} \frac{1/\lambda_i}{((1-\epsilon)\mu_1 - \mu_i)^2}, \max_{e_j} \frac{1/\lambda_j}{\left(\left(\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon}\right) - \mu_j\right)^2} \right\}$$

where the final step follows since $\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon} > \mu_i \forall i$ and $\mu_j \leq \mu_1$. The third inequality follows by the same approach as taken for $H^{\texttt{MILK1}}$ of multiplying the denominator by $(1-\epsilon)^2/(1-\epsilon)^2$.

Hence, we have that

$$H(\theta_*) \geq \min_\lambda \max_i \max\left\{ \frac{\frac{1}{\lambda_i}}{((1-\epsilon)\mu_1 - \mu_i)^2}, \frac{\frac{1}{\lambda_i}}{(\mu_1 + \frac{\tilde{\alpha}_\epsilon}{1-\epsilon} - \mu_i)^2}, \frac{\frac{1}{\lambda_i}}{(\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon} - \mu_i)^2} \right\}.$$

Solving for λ gives

$$H(\theta_*) \geq \sum_{i=1}^n \max\left\{ \frac{1}{((1-\epsilon)\mu_1 - \mu_i)^2}, \frac{1}{(\mu_1 + \frac{\tilde{\alpha}_\epsilon}{1-\epsilon} - \mu_i)^2}, \frac{1}{(\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon} - \mu_i)^2} \right\} = c_1 \cdot H_{(ST)^2}$$

for a constant c_1. To upper bound $H^{\texttt{MILK}}(\theta_*)$, we may choose a specific λ. Choosing

$$\lambda_i := \frac{\max\{((1-\epsilon)\mu_1 - \mu_i)^{-2}, (\mu_1 + \frac{\tilde{\alpha}_\epsilon}{1-\epsilon} - \mu_i)^{-2}, (\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon} - \mu_i)^{-2}\}}{\sum_j \max\{((1-\epsilon)\mu_1 - \mu_j)^{-2}, (\mu_1 + \frac{\tilde{\alpha}_\epsilon}{1-\epsilon} - \mu_j)^{-2}, (\mu_1 + \frac{\tilde{\beta}_\epsilon}{1-\epsilon} - \mu_j)^{-2}\}},$$

a similar computation shows that $H^{\texttt{MILK}}(\theta_*) \leq c_2 H_{(ST)^2}$ for a constant c_2. □

B.4.3 Upper Bound

First we restate Theorem 9 bounding the sample complexity of MILK.

Theorem 22. *Fix $\delta > 0$, threshold $\alpha > 0$, tolerance $\widetilde{\beta}$, and regularization $\gamma > 0$. Define the quantities $\Delta_{\min}^{\textit{Above}}(\epsilon) = \min_{\boldsymbol{x} \in G_\epsilon} \min_{\boldsymbol{x}'} \theta_*^\top(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))$ and $\Delta_{\min}^{\textit{Below}}(\epsilon) = \min_{\boldsymbol{x} \in G_\epsilon^c} \max_{\boldsymbol{x}' : (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^\top \theta_* < 0} (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^\top \theta_*$, and $\Delta_{\min} = \min\left\{\Delta_{\min}^{\textit{Above}}(\epsilon), \Delta_{\min}^{\textit{Below}}(\epsilon)\right\}$. Define also*

$$\bar{\beta}(\epsilon) = \min\{\beta > 0 : 4(\sqrt{\gamma}\|\theta_*\| + h)(2 + \sqrt{f(\mathcal{X}, \{\boldsymbol{y} \in \mathcal{Y}^\epsilon(\mathcal{X} \times \mathcal{X}) : |\boldsymbol{y}^\top \theta_*| \leq \beta\}; \gamma)}) \leq \beta\}.$$

With probability $1 - \delta$, MILK returns a set $\widehat{\mathcal{R}} = (\mathcal{X} \setminus \widehat{B}_t)$ at a time T_δ such that

$$\{\boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}) \geq (1-\epsilon) f(\boldsymbol{x}_*) + \bar{\beta}(\epsilon)\} \subseteq \widehat{\mathcal{R}} \subseteq \{\boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}) \geq (1-\epsilon) f(\boldsymbol{x}_*) - \widetilde{\beta} - \bar{\beta}(\epsilon)\}$$

and for any α, $\widetilde{\beta}$ *such that* $\max(\Delta_{\min}(\epsilon), \widetilde{\beta}) \geq \bar{\beta}(\epsilon)$

$$T_\delta \leq 256(B^2+\sigma^2)H^{MILK}(\theta_*)\log\left(\frac{4|\mathcal{X}|^2\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil^2}{\delta}\right) + 2\log\left(\frac{|\mathcal{X}|}{\delta}\right)\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil$$

for a sufficiently large constant c *where* $H^{MILK}(\theta_*) = \min_{\lambda\in\triangle_\mathcal{X}}\max\left\{H_\lambda^{MILK1}(\theta_*), H_\lambda^{MILK2}(\theta_*)\right\}$ *and*

$$H_\lambda^{MILK1}(\theta_*) := \max_{\boldsymbol{x}\in G_\epsilon}\max_{\boldsymbol{x}'\in\mathcal{X}}\frac{\|\phi(\boldsymbol{x})-(1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{((\phi(\boldsymbol{x})-(1-\epsilon)\phi(\boldsymbol{x}'))^\top\theta_*)^2, \widetilde{\beta}^2\}}$$

$$H_\lambda^{MILK2}(\theta_*) := \max_{\boldsymbol{x}\in G_\epsilon^c}\max_{\boldsymbol{x}'}\frac{\|\phi(\boldsymbol{x})-(1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{((\phi(\boldsymbol{x})-(1-\epsilon)\phi(\boldsymbol{x}_*))^\top\theta_*)^2, \widetilde{\beta}^2\}}.$$

Now we show a high probability concentration result that we will use for the remainder of this section.

Lemma 27. *For any* $\mathcal{V}\subset\mathcal{Y}^\epsilon(\mathcal{X}\times\mathcal{X})$ *define* $f(\mathcal{X},\mathcal{V};\gamma) = \min_{\lambda\in\triangle_\mathcal{X}}\max_{\boldsymbol{v}\in\mathcal{V}}\|\boldsymbol{v}\|^2_{(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I)^{-1}}$. *In each round* t, *define the event*

$$\mathcal{E}_t = \{|\boldsymbol{y}^T(\widehat{\theta}_t-\theta_*)| \leq 2^{-t} + (\sqrt{\gamma}\|\theta_*\|+h)\sqrt{f(\mathcal{X},\mathcal{Y}^\epsilon(\mathcal{Y}^\epsilon(\mathcal{A}_t));\gamma)}\;\forall\,\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{A}_t)\}$$

Holds $\mathbb{P}(\bigcup_{t=1}^\infty\mathcal{E}_t^c)\leq\delta$.

Proof. Using Theorem 6, for any $\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{A}_t)$ we have that with probability at least $1-\delta_t/|\mathcal{X}|^2$

$$\begin{aligned}
|\boldsymbol{y}^T(\widehat{\theta}_t-\theta_*)| &\leq \|\boldsymbol{y}\|_{\left(\sum_{\boldsymbol{x}\in\mathcal{X}}\lambda_{\boldsymbol{x}}\phi(\boldsymbol{x})\phi(\boldsymbol{x})^T+\gamma I\right)^{-1}}\left(\sqrt{\gamma}\|\theta_*\|+h+c\sqrt{\tfrac{(B^2+\sigma^2)}{N_t}\log(2t^2|\mathcal{X}|^2/\delta)}\right)\\
&\leq \sqrt{f(\mathcal{X},\mathcal{Y}^\epsilon(\mathcal{A}_t);\gamma)}\left(\sqrt{\gamma}\|\theta_*\|+h+2^{-t}/\sqrt{f(\mathcal{X},\mathcal{Y}^\epsilon(\mathcal{A}_t);\gamma)}\right)\\
&\leq 2^{-t}+(\sqrt{\gamma}\|\theta_*\|+h)\sqrt{f(\mathcal{X},\mathcal{Y}^\epsilon(\mathcal{A}_t);\gamma)}
\end{aligned}$$

Since $|\mathcal{Y}^\epsilon(\mathcal{A}_t)|\leq|\mathcal{X}|^2$, $\mathcal{E}_t$ holds for all $\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{A}_t)$ with probability $1-\delta_t$ via a union bound. Taking a second union bound over rounds, we have that

$$\mathbb{P}\left(\bigcup_{t=1}^\infty\mathcal{E}_t^c\right)\leq\sum_{t=1}^\infty\mathbb{P}(\mathcal{E}_t^c)\leq\sum_{t=1}^\infty\delta_t=\sum_{t=1}^\infty\frac{\delta}{2t^2}\leq\delta$$

□

Define

$$\bar{t} = \max\{t : (\sqrt{\gamma}\|\theta_*\|_2+h)(2+\sqrt{f(\mathcal{X},\{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{X}\times\mathcal{X}) : |\boldsymbol{y}^T\theta_*|\leq 2^{-t+2}\};\gamma)})\leq 2^{-t}\}.$$

As we will see in Lemmas 30 and 31,

$$\mathcal{Y}^\epsilon(\mathcal{A}_t)\subset\left\{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{X}\times\mathcal{X}) : |\boldsymbol{y}^T\theta_*|\leq 2^{-t+1}\right\}.$$

Thus for $t\leq\bar{t}$, holds on $\bigcap_t\mathcal{E}_t$ that

$$\forall\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{A}_t)\;,\;|\boldsymbol{y}^T(\widehat{\theta}_t-\theta_*)|\leq 2\cdot 2^{-t}.$$

Lemma 28. *On $\bigcap_t \mathcal{E}_t$, when $t \le \bar{t}$ holds $\widehat{G}_t \subset G_\epsilon^\phi := \{\boldsymbol{x} \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* > 0 \,\forall\, \boldsymbol{x}' \in \mathcal{X}\}$.*

Proof.

$$\begin{aligned}
\boldsymbol{x} \in \widehat{G}_t &\iff \forall \boldsymbol{x}' \,\exists t_{x'} \le \bar{t} \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\widehat{\theta}_{t_{x'}} \ge 2\cdot 2^{-t_{x'}} \\
&\iff \forall \boldsymbol{x}' \,\exists t_{x'} \le \bar{t} \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T(\widehat{\theta}_{t_{x'}} - \theta_*) + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* \ge 2\cdot 2^{-t_{x'}} \\
&\overset{\bigcap_t \mathcal{E}_t}{\implies} \forall \boldsymbol{x}' \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* > 0 \\
&\iff \boldsymbol{x} \in G_\epsilon^\phi.
\end{aligned}$$

□

Lemma 29. *On $\bigcap_t \mathcal{E}_t$, when $t \le \bar{t}$ holds $\widehat{B}_t \subset (G_\epsilon^\phi)^c$.*

Proof.

$$\begin{aligned}
\boldsymbol{x} \in \widehat{B}_t &\iff \exists \boldsymbol{x}', t_{x'} \le \bar{t} \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\widehat{\theta}_t \le -2\cdot 2^{-t_{x'}} \\
&\iff \exists \boldsymbol{x}', t_{x'} \le \bar{t} \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T(\widehat{\theta}_t - \theta_*) + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* \le -2\cdot 2^{-t_{x'}} \\
&\overset{\bigcap_t \mathcal{E}_t}{\implies} \exists \boldsymbol{x}' \,:\, (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* > \epsilon \\
&\iff \boldsymbol{x} \in (G_\epsilon^\phi)^c.
\end{aligned}$$

□

Lemma 30. *On the event $\bigcap_t \mathcal{E}_t$ for $t \le \bar{t}$,*

$$\{(\boldsymbol{x}, \boldsymbol{x}') : (\boldsymbol{x}, \boldsymbol{x}'), \boldsymbol{x} \in G_\epsilon^\phi\} \subset \left\{(\boldsymbol{x}, \boldsymbol{x}') \middle| |(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_*| \le 2^{-t+2}\right\} =: \mathcal{S}_t^{\textit{Above}}$$

Proof. On $\bigcap_t \mathcal{E}_t$ for $t \le \bar{t}$, for any $\boldsymbol{y} \in \mathcal{A}_t$

$$|\boldsymbol{y}^T\widehat{\theta}_t| \ge |\boldsymbol{y}^T\theta_*| - |\boldsymbol{y}^T(\widehat{\theta}_t - \theta_*)| \overset{\mathcal{E}_t}{\ge} |\boldsymbol{y}^T\theta_*| - 2\cdot 2^{-t}.$$

For $\boldsymbol{y}$ such that $|\boldsymbol{y}^T\theta_*| \ge 2\cdot 2^{-t+1}$, the above implies that

$$|\boldsymbol{y}^T\widehat{\theta}_t| \ge 2\cdot 2^{-t}.$$

By the elimination condition, this implies that $\boldsymbol{y}$ is removed from $\mathcal{A}_t$. Hence

$$\mathcal{A}_{t+1} \subset \left\{\boldsymbol{y} \in \mathcal{Y}(\mathcal{X}) : |\boldsymbol{y}^T\theta_* - \epsilon| \le 2\cdot 2^{-t+1}\right\}.$$

Specializing this argument to $\{(\boldsymbol{x}, \boldsymbol{x}') : (\boldsymbol{x}, \boldsymbol{x}'), \boldsymbol{x} \in G_\epsilon^\phi\} \subset \mathcal{A}_t$ completes the proof. □

Lemma 31. *On the event $\bigcap_t \mathcal{E}_t$ for $t \le \bar{t}$,*

$$\begin{aligned}
\{(\boldsymbol{x}, \boldsymbol{x}') : (\boldsymbol{x}, \boldsymbol{x}'), \boldsymbol{x} \in (G_\epsilon^\phi)^c\} \subset \Bigg\{(\boldsymbol{x}, \boldsymbol{x}') \Bigg| &|(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* - \epsilon| \le 2^{-t+2} \\
&\textit{and } \{(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_*\} \ge -2^{-t+2}\Bigg\} =: \mathcal{S}_t^{\textit{Below}}
\end{aligned}$$

Proof. The guarantee that $|(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* - \epsilon| \le 2^{-t+2}$ for any $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{A}_t$ follows by the same argument as Lemma 30. For the additional statement, that $(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_* \ge -2^{-t+2}$, note that if

$$(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\widehat{\theta}_t \le -2^{-t+1}$$

then the pair $(\boldsymbol{x}, \boldsymbol{x}_*)$ is eliminated from $\mathcal{A}_t$. If

$$(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\widehat{\theta}_t \le -2^{-t+2},$$

then using this and the event $\bigcap_t \mathcal{E}_t$

$$(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\widehat{\theta}_t = (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T(\widehat{\theta}_t - \theta_*) + (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_* \le -2^{-t+1}.$$

Hence, the only pairs $(\boldsymbol{x}, \boldsymbol{x}_*)$ that remain in $\mathcal{A}_t$ where $\boldsymbol{x}_* \in (G_\epsilon^\phi)^c$ are such that $(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_*\} \ge -2^{-t+2}$. We conclude by noting that the above argument for $\boldsymbol{x}_*$ could be repeated for any $\boldsymbol{x}'$ such that $(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* < 0$. □

Remark: Lemmas 30 and 31 jointly imply that $\mathcal{A}_t \subset \mathcal{S}_t^{\text{Above}} \cup \mathcal{S}_t^{\text{Below}} =: \mathcal{S}_t$ for $t \le \bar{t}$. Furthermore, $f(\mathcal{X}, \mathcal{Y}^\epsilon(\mathcal{A}_t), \gamma) \le f(\mathcal{X}, \mathcal{Y}^\epsilon(\mathcal{S}_t), \gamma)$.

Remark:
The algorithm stops on either of two conditions. On one hand if $t \ge \lceil \log_2(4/\widetilde{\beta}) \rceil =: t_\beta$, then it has achieved precision $\widetilde{\beta}$ as desired and it terminates. Otherwise, it terminates if $\widehat{G}_t \cup \widehat{B}_t = \mathcal{X}$. This occurs when $\widetilde{\beta}$ is very small. Define the quantities $\Delta_{\min}^{\text{Above}}(\epsilon) = \min_{\boldsymbol{x} \in G_\epsilon} \min_{\boldsymbol{x}'} \theta_*^\top(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))$ and $\Delta_{\min}^{\text{Below}}(\epsilon) = \min_{\boldsymbol{x} \in G_\epsilon^c} \max_{\boldsymbol{x}' : (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^\top\theta_* < 0} (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^\top\theta_*$, and $\Delta_{\min}(\epsilon) = \min\left\{\Delta_{\min}^{\text{Above}}(\epsilon), \Delta_{\min}^{\text{Below}}(\epsilon)\right\}$. Recall

$$\begin{aligned}
\bar{t} &= \max\{t : (\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{\boldsymbol{y} \in \mathcal{Y}^\epsilon(\mathcal{X} \times \mathcal{X}) : |\boldsymbol{y}^T\theta_*| \le 4 \cdot 2^{-t}\}, \gamma)}) \le 2^{-t}\} \\
&= \max\{t : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{\boldsymbol{y} \in \mathcal{Y}^\epsilon(\mathcal{X} \times \mathcal{X}) : |\boldsymbol{y}^T\theta_*| \le 4 \cdot 2^{-t}\}, \gamma)}) \le 4 \cdot 2^{-t}\} \\
&= -2 + \max\{t : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{\boldsymbol{y} \in \mathcal{Y}^\epsilon(\mathcal{X} \times \mathcal{X}) : |\boldsymbol{y}^T\theta_*| \le 2^{-t}\}, \gamma)}) \le 2^{-t}\} \\
&= -3 + \log_2(\min\{\beta > 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{\boldsymbol{y} \in \mathcal{Y}^\epsilon(\mathcal{X} \times \mathcal{X}) : |\boldsymbol{y}^T\theta_*| \le \beta\}, \gamma)}) \le \beta\}).
\end{aligned}$$

This defines

$$\bar{\beta}(\epsilon) = \min\{\beta > 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2 + \sqrt{f(\mathcal{X}, \{\boldsymbol{y} \in \mathcal{Y}^\epsilon(\mathcal{X} \times \mathcal{X}) : |\boldsymbol{y}^T\theta_*| \le \beta\}, \gamma)}) \le \beta\}.$$

Let $t_{\max}$ denote the random variable of the last round before the algorithm terminates. The following Lemmas give a guarantee on the set $\mathcal{X} \setminus \widehat{B}_t$ at termination.

Lemma 32. *On the event $\bigcap_{t=1}^\infty \mathcal{E}_t$, `MILK` returns a set $(\mathcal{X} \setminus \widehat{B}_{t_{\max}})$ such that $\{\boldsymbol{x} : f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\epsilon)\} \subset (\mathcal{X} \setminus \widehat{B}_{t_{\max}})$.*

Proof. Take any $\boldsymbol{x}$ such that $f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\alpha)$ and recall that by assumption $|f(\boldsymbol{x}) - \phi(\boldsymbol{x})^T\theta_*| \le h$ for all $\boldsymbol{x} \in \mathcal{X}$. We consider two cases. In the first case, assume that $t_{\max} \le \bar{t}$. We claim that in this case $\nexists t$ such that $\boldsymbol{x} \in \widehat{B}_t$. We prove this by contradiction. Assume not. Then $\exists t$ and a $\boldsymbol{x}'$ such that

$$\widehat{\theta}_t^T(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')) < -2^{-t+1}$$

$$
\begin{aligned}
&\iff (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T(\widehat{\theta}_t - \theta_*) + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* < -2^{-t+1} \\
&\overset{\mathcal{E}_t, t_{\max} \leq \bar{t}}{\implies} -2^{-t+1} + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))\theta_* < -2^{-t+1} \\
&\iff (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))\theta_* < 0 \\
&\implies f(\boldsymbol{x}) - (1-\epsilon)f(\boldsymbol{x}') < h + (1-\epsilon)h
\end{aligned}
$$

Recall that we have assumed that $f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\alpha)$ and $\bar{\beta}(\epsilon) > 4h$ by definition. Hence, this implies that

$$
(1-\epsilon)f(\boldsymbol{x}_*) - (1-\epsilon)f(\boldsymbol{x}') < h + (1-\epsilon)h - \bar{\beta}(\alpha) < 0
$$

which is a contradiction since $f(\boldsymbol{x}_*) \geq f(\boldsymbol{x}')$ by definition. Hence, we have shown in the case that $t_{\max} \leq \bar{t}$, $\{\boldsymbol{x} : f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\epsilon)\} \subset (\mathcal{X} \setminus \widehat{B}_{t_{\max}})$.

In the second case, assume that $t_{\max} > \bar{t}$ and take $\boldsymbol{x}$ such that $f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\alpha)$. We claim that $\boldsymbol{x} \in \widehat{G}_{\bar{t}}$ and hence $(\boldsymbol{x}, \boldsymbol{x}') \notin \mathcal{A}_t$ for any $t > \bar{t}$ and thus $\boldsymbol{x}$ is never added to $\widehat{B}_t$. This occurs if for every $\phi(\boldsymbol{x}')$

$$
\begin{aligned}
&(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\widehat{\theta}_{\bar{t}} > 2^{-\bar{t}+1} \\
&\iff (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T(\widehat{\theta}_{\bar{t}} - \theta_*) + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* > 2^{-\bar{t}+1} \\
&\overset{\mathcal{E}_{\bar{t}}}{\impliedby} -2^{-\bar{t}+1} + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* \geq 2^{-\bar{t}+1} \\
&\iff (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* \geq 2^{-\bar{t}+2} \\
&\iff (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_* \geq 0.5\bar{\beta}(\epsilon) \\
&\impliedby f(\boldsymbol{x}) - (1-\epsilon)f(\boldsymbol{x}') \geq 0.5\bar{\beta}(\epsilon) + h + (1-\epsilon)h
\end{aligned}
$$

where the penultimate step follows by definition of $\bar{\beta}(\epsilon)$. Recall that $f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\alpha)$. Hence, the above is implied by

$$
\begin{aligned}
&(1-\epsilon)f(\boldsymbol{x}_*) + \bar{\beta}(\alpha) - (1-\epsilon)f(\boldsymbol{x}') \geq 0.5\bar{\beta}(\epsilon) + h + (1-\epsilon)h \\
&\impliedby \bar{\beta}(\epsilon) \geq 0.5\bar{\beta}(\epsilon) + h + (1-\epsilon)h
\end{aligned}
$$

where the final step follows by noting that $f(\boldsymbol{x}_*) \geq f(\boldsymbol{x}')$ for any $\boldsymbol{x}'$. The final statement is true since $\bar{\beta}(\epsilon)$ and thus implies the claim. Therefore, we have shown that $\boldsymbol{x} \in \widehat{G}_{\bar{t}}$ and is therefore not added to $\widehat{B}_t$ in a later round. These two cases together complete the proof. □

Lemma 33. *On the event $\bigcap_{t=1}^{\infty} \mathcal{E}_t$, `MILK` returns a set $(\mathcal{X} \setminus \widehat{B}_{t_{\max}})$ such that $(\mathcal{X} \setminus \widehat{B}_{t_{\max}}) \subset \{\boldsymbol{x} : f(\boldsymbol{x}) > (1-\epsilon)f(\boldsymbol{x}_*) - \bar{\beta}(\epsilon) - \widetilde{\beta}\}$.*

Proof. Take any $\boldsymbol{x}$ such that $f(\boldsymbol{x}) < (1-\epsilon)f(\boldsymbol{x}_*) - \bar{\beta}(\epsilon) - \widetilde{\beta}$. We claim that there exists a $t \leq t_{\max}$ such that $\boldsymbol{x}$ is added to $\widehat{B}_t$ which implies that $\boldsymbol{x} \notin (\mathcal{X} \setminus \widehat{B}_{t_{\max}})$. Suppose for contradiction that this is not the case. Then for all $t \leq t_{\max}$,

$$
\begin{aligned}
&\widehat{\theta}_t^T(\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*)) > -2^{-t+1} \\
&\iff (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T(\widehat{\theta}_t - \theta_*) + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_* > -2^{-t+1} \\
&\overset{\mathcal{E}_t}{\implies} 2^{-t} + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)} + (\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_* > -2^{-t+1} \\
&\implies (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{A}_t; \gamma)} + f(\boldsymbol{x}) - (1-\epsilon)f(\boldsymbol{x}_*) + h + (1-\epsilon)h > -2^{-t+1} - 2^{-t}
\end{aligned}
$$

$$\implies f(\boldsymbol{x}) - (1-\epsilon)f(\boldsymbol{x}_*) > -2^{-t+1} - 2^{-t} - h - (1-\epsilon)h - (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}.$$

Plugging in $f(\boldsymbol{x}) < (1-\epsilon)f(\boldsymbol{x}_*) - \bar{\beta}(\epsilon) - \widetilde{\beta}$, the above implies

$$\bar{\beta}(\epsilon) + \widetilde{\beta} < 2^{-t+1} + 2^{-t} + h + (1-\epsilon)h + (\sqrt{\gamma}\|\theta_*\| + h)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)} \tag{B.2}$$

Next, recall that MILK terminates either on the condition that $t = \lceil \log_2(4/\widetilde{\beta}) \rceil$ or that $\widehat{G}_t \cup \widehat{B}_t = \mathcal{X}$. Using this, we brake our analysis into cases.

Case 1: $t_{\max} = \lceil \log_2(4/\widetilde{\beta}) \rceil \leq \bar{t}$.

In this case, MILK stops due to the $\widetilde{\beta}$ tolerance in a round before $\bar{t}$. For $t \leq \bar{t}$, we have that $2^{-t} \geq + \left(\sqrt{\gamma}\|\theta_*\| + h\right)\sqrt{f(\mathcal{X}, \mathcal{S}_t; \gamma)}$. Hence, the above implies that

$$\bar{\beta}(\alpha) + \widetilde{\beta} < 2^{-t+2} + h + (1-\epsilon)h.$$

As we have assumed this condition for all $t \leq t_{\max}$, we may plug in $t_{\max}$ which implies

$$\bar{\beta}(\alpha) + \widetilde{\beta} < \widetilde{\beta} + h + (1-\epsilon)h.$$

As $\bar{\beta}(\alpha) > 4h$, this is a contradiction. Hence there must exist a t such that $\boldsymbol{x} \in \widehat{B}_t$.

Case 2: $t_{\max} \leq \bar{t} < \lceil \log_2(4/\widetilde{\beta}) \rceil$.

In this case, MILK terminates before round $t = \lceil \log_2(4/\widetilde{\beta}) \rceil$. Hence, it does so on the condition that $\widehat{G}_t \cup \widehat{B}_t = \mathcal{X}$. Note that for $f(\boldsymbol{x}) < \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}$, we have that $\boldsymbol{x} \in (G_\alpha^\phi)^c$ since $\bar{\beta}(\alpha) > h$ and $\widetilde{\beta} \geq 0$. If we terminate before round $\bar{t}$, we have by Lemma 29 that $(G_\alpha^\phi)^c \subset \widehat{B}_t$ which implies that $\boldsymbol{x} \in \widehat{B}_{t_{\max}}$. This contradicts the assumption that $\nexists t : \boldsymbol{x} \in \widehat{B}_t$.

Case 3: $\bar{t} < t_{\max}$.

In this case, MILK terminates at a round after $\bar{t}$. In this setting, we argue that $\boldsymbol{x} \in \widehat{B}_{\bar{t}}$. Recall that for any $t \leq \bar{t}$, (B.2) simplifies to

$$\bar{\beta}(\alpha) + \widetilde{\beta} < 2^{-t+2} + h + (1-\epsilon)h.$$

Plugging in $\bar{t}$, and noting that $2^{-\bar{t}+2} = \frac{1}{2}\bar{\beta}(\alpha)$, the above implies

$$\bar{\beta}(\alpha) + \widetilde{\beta} < \frac{1}{2}\bar{\beta}(\alpha) + h + (1-\epsilon)h.$$

Noting that $\bar{\beta}(\alpha) > 4h$, shows that the above is a contradiction. Hence, there exists a $t \leq \bar{t}$ such that $\boldsymbol{x} \in \widehat{B}_t$.

Therefore, in all cases we have shown that for any $\boldsymbol{x}$ such that $f(\boldsymbol{x}) < \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}$, $\boldsymbol{x} \in \widehat{B}_t$. Therefore, for the returned set $\mathcal{X} \setminus \widehat{B}_{t_{\max}}$, we have that

$$(\mathcal{X} \setminus \widehat{B}_{t_{\max}}) \subset \{\boldsymbol{x} : f(\boldsymbol{x}) > \alpha - \bar{\beta}(\alpha) - \widetilde{\beta}\}.$$

□

Proof of Theorem 9. Throughout, assume the high probability event $\bigcap_T \mathcal{E}_t$. By Lemmas 32 and 33 in conjunction with the high probability event $\bigcap \mathcal{E}_t$ we have correctness. It remains to control the sample complexity of MILK. Recall that we have assumed that $\max(\Delta_{\min}(\epsilon), \widetilde{\beta}) \geq \bar{\beta}(\epsilon)$. This implies that

$\min\{\lceil\log_2(4/\Delta_{\min}(\epsilon))\rceil, \lceil\log_2(4/\widetilde{\beta})\rceil\} \leq \bar{t}$. Applying Lemmas 30 and 31, we have that $t_{\max} \leq \min\{\lceil\log_2(4/\Delta_{\min}(\epsilon))\rceil, \lceil\log_2(4/\widetilde{\beta})\rceil\} \leq \bar{t}$ and that $\mathcal{A}_t \subseteq \mathcal{S}_t$ for all rounds t. Now we proceed by bounding the total number of samples drawn.

$$
\begin{aligned}
\tau &\leq \sum_{t=1}^{t_{\max}} N_t \\
&\leq \sum_{t=1}^{\min\{\lceil\log_2(4/\Delta_{\min}(\epsilon))\rceil, \lceil\log_2(4/\widetilde{\beta})\rceil\}} N_t \\
&= \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} N_t \\
&= \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} \max\left\{c_1 \log(|\mathcal{X}|/\delta), c^2 2^{2t} f(\mathcal{Y}^\epsilon(\mathcal{A}_t);\gamma)(B^2+\sigma^2)\log(2t^2|\mathcal{X}|^2/\delta)\right\} \\
&\leq c_1 \log(|\mathcal{X}|/\delta)\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil + c^2(B^2+\sigma^2) \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} 2^{2t} f(\mathcal{Y}^\epsilon(\mathcal{A}_t);\gamma)\cdot\log(2t^2|\mathcal{X}|^2/\delta) \\
&= c_1 \log(|\mathcal{X}|/\delta)\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil + \\
&\qquad c^2(B^2+\sigma^2) \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} 2^{2t} \min_{\lambda\in\triangle_{\mathcal{X}}} \max_{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{A}_t)} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}} \cdot \log(2t^2|\mathcal{X}|^2/\delta) \\
&\leq c_1 \log(|\mathcal{X}|/\delta)\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil + \\
&\qquad c^2(B^2+\sigma^2)\log\left(\frac{4|\mathcal{X}|^2\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil^2}{\delta}\right) \\
&\qquad \cdot \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} 2^{2t} \min_{\lambda\in\triangle_{\mathcal{X}}} \max_{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{A}_t)} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}} \\
&\leq c_1 \log(|\mathcal{X}|/\delta)\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil + \\
&\qquad c^2(B^2+\sigma^2)\log\left(\frac{4|\mathcal{X}|^2\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil^2}{\delta}\right) \\
&\qquad \cdot \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} 2^{2t} \min_{\lambda\in\triangle_{\mathcal{X}}} \max_{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{S}_t)} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}} \\
&= c_1 \log(|\mathcal{X}|/\delta)\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil + \\
&\qquad c^2(B^2+\sigma^2)\log\left(\frac{4|\mathcal{X}|^2\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil^2}{\delta}\right) \\
&\qquad \cdot \sum_{t=1}^{\lceil\log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} \min_{\lambda\in\triangle_{\mathcal{X}}} \max\left\{2^{2t} \max_{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{S}_t^{\text{Above}})} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}}, 2^{2t} \max_{\boldsymbol{y}\in\mathcal{Y}^\epsilon(\mathcal{S}_t^{\text{Below}})} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}}\right\}.
\end{aligned}
$$

where the final equality follows by partitioning $\mathcal{S}_t = \mathcal{S}_t^{\text{Above}} \cup \mathcal{S}_t^{\text{Below}}$.

Focusing on this final summation, note that

$$\frac{1}{\lceil \log_2(4(\Delta_{\min}(\epsilon) \vee \widetilde{\beta})^{-1})\rceil} \sum_{t=1}^{\lceil \log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} 2^{2t} \min_{\lambda\in\triangle_{\mathcal{X}}} \max\left\{ \max_{\boldsymbol{y}\in\mathcal{S}_t^{\text{Above}}} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}}, \max_{\boldsymbol{y}\in\mathcal{S}_t^{\text{Below}}} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}} \right\}$$

$$\leq \max_{t\leq \lceil \log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} \min_{\lambda\in\triangle_{\mathcal{X}}} 2^{2t} \max\left\{ \max_{\boldsymbol{y}\in\mathcal{S}_t^{\text{Above}}} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}}, \max_{\boldsymbol{y}\in\mathcal{S}_t^{\text{Below}}} \|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}} \right\}$$

$$\leq \min_{\lambda\in\triangle_{\mathcal{X}}} \max_{t\leq \lceil \log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} \max\left\{ \max_{\boldsymbol{y}\in\mathcal{S}_t^{\text{Above}}} 2^{2t}\|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}}, \max_{\boldsymbol{y}\in\mathcal{S}_t^{\text{Below}}} 2^{2t}\|\boldsymbol{y}\|^2_{(A(\lambda)+\gamma I)^{-1}} \right\}$$

$$= \min_{\lambda\in\triangle_{\mathcal{X}}} \max_{t\leq \lceil \log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} \max\left\{ \max_{(\boldsymbol{x},\boldsymbol{x}')\in\mathcal{S}_t^{\text{Above}}} 2^{2t}\|\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}} , \right.$$

$$\left. \max_{(\boldsymbol{x},\boldsymbol{x}')\in\mathcal{S}_t^{\text{Below}}} 2^{2t}\|\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}} \right\}$$

$$\overset{\text{Lemmas 30, 31, }\widetilde{\beta}}{\leq} 16 \min_{\lambda\in\triangle_{\mathcal{X}}} \max_{t\leq \lceil \log_2(4(\Delta_{\min}(\epsilon)\vee\widetilde{\beta})^{-1})\rceil} \max\left\{ \max_{(\boldsymbol{x},\boldsymbol{x}')\in\mathcal{S}_t^{\text{Above}}} \frac{\|\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{((\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_*)^2, \widetilde{\beta}^2\}}, \right.$$

$$\left. \max_{(\boldsymbol{x},\boldsymbol{x}')\in\mathcal{S}_t^{\text{Below}}} \frac{\|\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{((\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_* - \epsilon)^2, \widetilde{\beta}^2\}} \right\}$$

$$\leq 16 \min_{\lambda\in\triangle_{\mathcal{X}}} \max\left\{ \max_{\boldsymbol{x}\in G_\epsilon} \max_{\boldsymbol{x}'} \frac{\|\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{((\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}'))^T\theta_*)^2, \widetilde{\beta}^2\}}, \right.$$

$$\left. \max_{\boldsymbol{x}\in G_\epsilon^c} \max_{\boldsymbol{x}'} \frac{\|\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}')\|^2_{(A(\lambda)+\gamma I)^{-1}}}{\max\{((\phi(\boldsymbol{x}) - (1-\epsilon)\phi(\boldsymbol{x}_*))^T\theta_* - \epsilon)^2, \widetilde{\beta}^2\}} \right\}$$

Plugging this in with $c = 4$ and $c_1 = 2$ from Theorem 20 for RIPS with the Catoni estimator completes the proof. □

B.5 Additional Experiment Details

In this section we discuss additional experimental details not covered in the main paper. We first give an overview of the algorithms implemented in the following section. All code was written in python and run on a 64 core cluster machine. We have included implementations of all methods and a demo file showing how to call and run the various algorithms.

B.5.1 Algorithms Implemented

In this section we briefly discuss the algorithms implemented and the hyper-parameters used in the algorithms. The algorithms implemented are s follows:

Gaussian Process Experiments For all the algorithms in this section we assumed a GP Prior $N(0, k(x, x'))$ where $k(x, x')$ was the RBF kernel given by $k(x, x') = \exp(-\|x - x'\|^2/2\ell^2)$.

At every time step we builds the confidence interval

$$Q_t(\boldsymbol{x}) \coloneqq \left[\mu_{t-1}(\boldsymbol{x}) \pm \beta_t^{1/2}\sigma_{t-1}(\boldsymbol{x})\right]$$

where μ_{t-1}, and σ_{t-1} is the posterior mean and variance function over the observed points. For an observation $\boldsymbol{y}_t$ at time t we define μ_{t-1}, and σ_{t-1} as follows:

$$\begin{aligned}
\mu_t(\boldsymbol{x}) &\coloneqq \boldsymbol{k}_t(\boldsymbol{x})^T \left(\boldsymbol{K}_t + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}_t \\
k_t\left(\boldsymbol{x}, \boldsymbol{x}'\right) &\coloneqq k\left(\boldsymbol{x}, \boldsymbol{x}'\right) - \boldsymbol{k}_t(\boldsymbol{x})^T \left(\boldsymbol{K}_t + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}_t(\boldsymbol{x}) \\
\sigma_t^2(\boldsymbol{x}) &\coloneqq k_t(\boldsymbol{x}, \boldsymbol{x})
\end{aligned}$$

where, $\boldsymbol{k}_t(\boldsymbol{x}) = \left[k\left(\boldsymbol{x}_1, \boldsymbol{x}\right), \ldots, k\left(\boldsymbol{x}_t, \boldsymbol{x}\right)\right]^T$ and $\boldsymbol{K}_t$ is the kernel matrix over the observed points.

1. **LSE:** We implemented the LSE algorithm by (Gotovos, 2013). This algorithm maintains an active set of unclassified points defined as U_t and the super-level set H_t and sub-level set L_t.

   At every round LSE selects the most ambiguous point, where the ambiguity is defined as

   $$a_t(\boldsymbol{x}) = \min\left\{\max\left(Q_t(\boldsymbol{x})\right) - \alpha, \alpha - \min\left(Q_t(\boldsymbol{x})\right)\right\}$$

   that is, the points LSE is most unsure to classify into H_t or L_t. Note that in contrast to this approach MELK follows the optimal allocation over the active set to select the next sample.

2. **TruVar:** We also implemented a modified version of TruVar(Bogunovic et al., 2016) with zero cost and homoscedastic noise. TruVar samples in such a fashion to ensure the maximum decrease of the posterior variance. As above, we maintain a Gaussian Process Posterior and we sample the arm

   $$\arg\max_{x \in \mathcal{X}} \sum_{\bar{x} \in \mathcal{A}_t} \sigma_t^2(\bar{x}) - \sum_{\bar{x} \in \mathcal{A}_t} \sigma_{t-1|x}^2(\bar{x})$$

   where $\sigma_{t-1|x}^2(\bar{x})$ is the posterior variance of $\bar{x}$ if we sample x.

3. **MELK:** As described in the text, we compute the means and variances of the arms using a Gaussian posterior (identical to above) and eliminate arms when their lower/upper bound is below/above the specified threshold τ. We implemented a batched sampling algorithm where we compute the design

   $$\min_{\lambda \in \mathcal{X}} \max_{z \in A_t} \|z\|^2_{(A(\lambda)+\gamma I)^{-2}}$$

   ever 10 samples and then sample from it. At the i-th calculation, $\gamma = 1/(10 * i)$. We also use the Frank-Wolfe method to compute the optimal allocation over the active set before every round as described in Section B.6. We set the step-size of Frank-Wolfe method as 1 and cap the maximum number of iteration to converge for Frank-Wolfe to 500.

Linear Bandits Examples

Additionally, we also consider comparing algorithms exactly as written using theoretically justified confidence widths in all cases. This presents a challenge as `MELK` and `MILK` are designed for the frequentist regime and `LSE` and `TruVar` are Bayesian in nature. To level the playing field, we consider all algorithms in the frequentist regime. For this experiment, we focused primarily on comparing `MELK` to `LSE` and `MILK` to `LSE-imp` `LSE` can naturally be adapted to the frequentist setting with the tight RKHS confidence bounds from (Chowdhury and Gopalan, 2017). These bounds scale with the maximum information gain Γ_T. To make the comparison fair, we consider all algorithms in the linear regime where $\Gamma_T = O(d \log(T))$. By contrast, for the squared exponential kernel, $\Gamma_T = O\left(\log(T)^d\right)$, and this leads to overly pessimistic

confidence widths preventing a meaningful comparison of the algorithms. Indeed, even for moderate d such as $d = 4$, LSE had confidence widths that were more that an order of magnitude wider for the squared exponential kernel. Hence, we focus on the case of the linear kernel for our experimental comparison where the differences are not so stark. Below, we describe all algorithms in this regime.

LSE follows the same acquisition function described in the previous section. We provide additional details about MELK, MILK, and LSE-imp in this setting.

1. **MELK:** We implement the MELK algorithm as defined in Algorithm 3.1. Recall that $|f(x)| \leq B$, and for the experiments we set $B = 1$. We set the confidence parameter $\delta = 0.1$, the regularization parameter $\gamma = 1e - 7$. Note that we use the original confidence width of $(B^2 + \sigma^2)\log(2t^2|\mathcal{X}|^2/\delta)$ as stated in our algorithm, where σ^2 is the noise parameter specific to the environment. We also use the Frank-Wolfe method to compute the optimal allocation over the active set before every round. We set the step-size of Frank-Wolfe method as 0.5 and cap the maximum number of iteration to converge for Frank-Wolfe to 2000.

2. **LSE-imp:** We implement the LSE-Implicit algorithm as stated in (Gotovos, 2013). LSE-Implicit proceeds quite similarly to LSE by constructing the confidence region $C_t(\boldsymbol{x})$ (as defined above) and classifying points to the sub-level set L_t or super-level set H_t. We set the confidence width as in LSE for calculating the confidence region. Note that LSE-Implicit works in the implicit level set estimation setting and so constructs an estimate of the function maximum to classify points into H_t or L_t. It builds an optimistic and pessimistic estimate of the function maximum as

$$f_t^{opt} \coloneqq \max_{x \in U_t} \max\left(C_t(x)\right), \quad f_t^{\text{pes}} = \max_{x \in U_t} \min\left(C_t(x)\right)$$

   respectively. A point $\boldsymbol{x}$ is classified into H_t if $\min\left(C_t(\boldsymbol{x})\right) \geq (1-\epsilon) f_t^{opt}$ or classified into L_t if $\max\left(C_t(\boldsymbol{x})\right) \leq (1-\epsilon) f_t^{pes}$. Finally, LSE-Implicit selects the next point with the largest confidence region width, defined as follows:

$$w_t(\boldsymbol{x}) = \max\left(C_t(\boldsymbol{x})\right) - \min\left(C_t(\boldsymbol{x})\right)$$

   such that this leads to more exploration. Again, note that in contrast MILK in Algorithm 3.2 uses the optimal allocation proportion over the active set to sample the next point.

3. **MILK:** We implement the MILK algorithm as stated in Algorithm 3.2. Note that MILK proceeds as similarly to MELK but with the allocation calculated over the difference of vectors $\mathcal{Y}^{\epsilon}(\mathcal{A})$ over the active set and a different elimination condition depending on ϵ. For MILK we set a similar hyper-parameters like MELK. We set the confidence parameter $\delta = 0.1$, the regularization parameter $\gamma = 1e - 7$, and the confidence width of $(B^2 + \sigma^2)\log(2t^2|\mathcal{X}|^2/\delta)$. We use the Frank-Wolfe method to compute the optimal allocation over the active set of points and set the step-size of Frank-Wolfe method as 0.5 and cap the maximum number of iteration to converge for Frank-Wolfe to 2000. Note that we set ϵ depending on specific environment setting.

B.5.2 Additional Experiments

All experiments were done with 25 repetitions. We consider the $f1$-scores on three environments considered below.

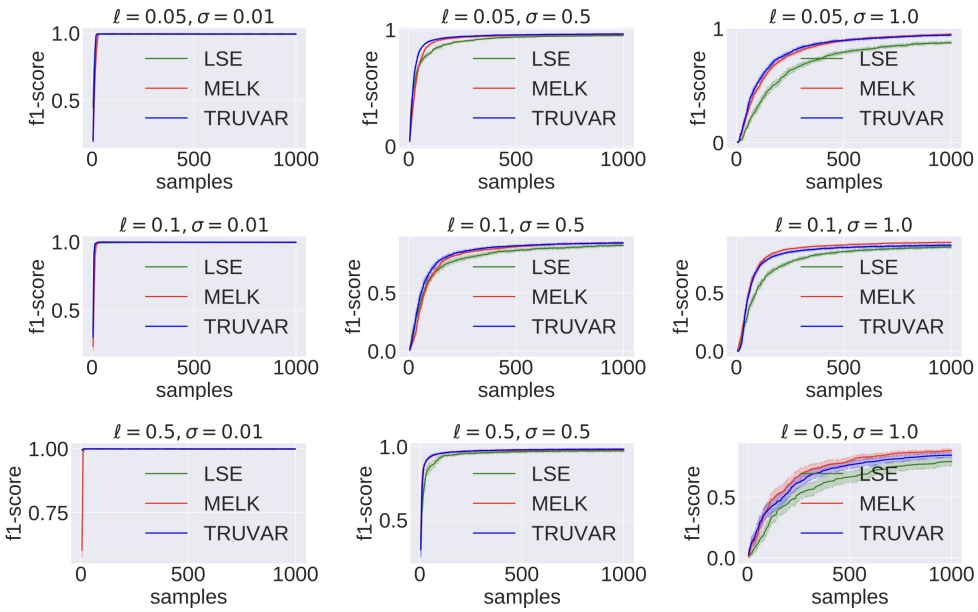


Figure B.1: f drawn randomly from a squared exponential kernel $N(0, k(\boldsymbol{x}, \boldsymbol{x}'))$. σ denotes the standard deviation of the noise and ℓ denotes the bandwidth of the kernel (i.e., $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|/2\ell^2)$).

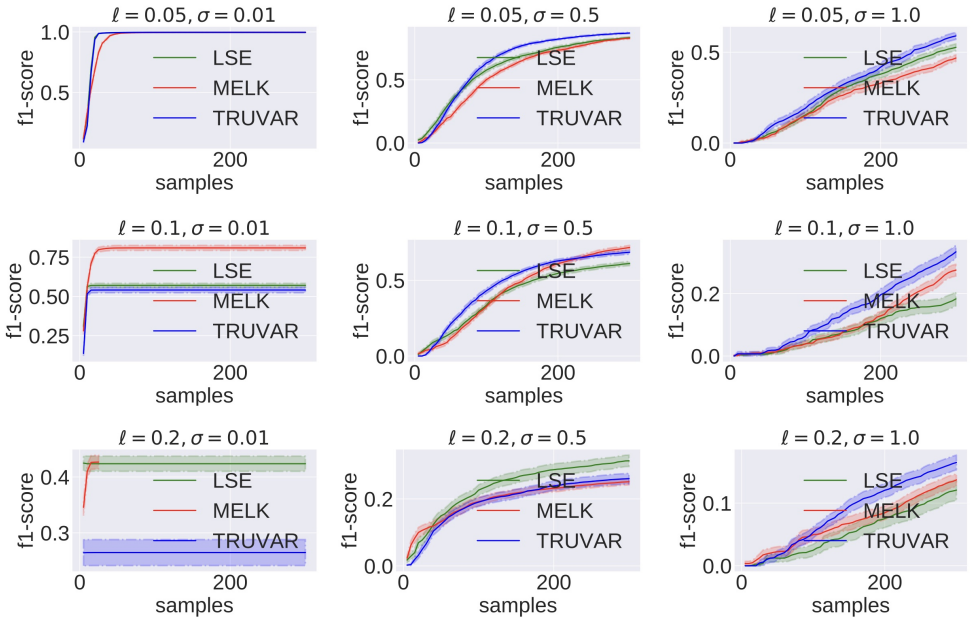


Figure B.2: $f(x) = \cos(8\pi x)$. σ denotes the standard deviation of the noise and ℓ denotes the bandwidth of the kernel (i.e., $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|/2\ell^2)$).

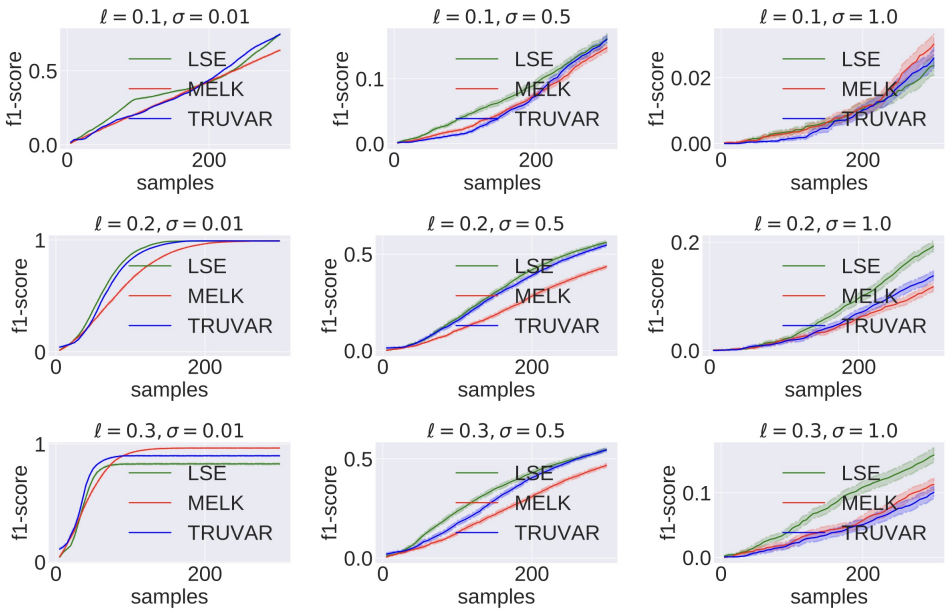
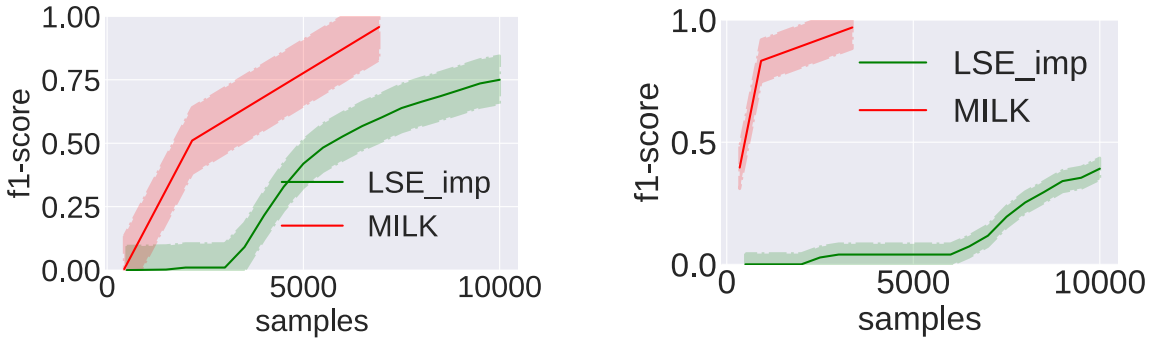


Figure B.3: $f(x, y) = \cos(2\pi x)\sin(2\pi y)$. σ denotes the standard deviation of the noise and ℓ denotes the bandwidth of the kernel (i.e., $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|/2\ell^2)$).

Linear Examples with true confidence widths

Finally we compare the performance of the methods using exact confidence widths.



(a) Linear, explicit

(b) Linear, Implicit

Figure B.4: Comparison of algorithms using theoretically justified confidence widths on a linear bandit setting.

For the Linear kernel experiments in Figures B.4a and B.4b, we run all algorithms with exact confidence intervals as specified by theoretical guarantees and use the theoretical upper bound on information gain γ_T shown in (Srinivas et al., 2009) for the confidence widths from (Valko et al., 2013) needed for LSE. We compare the methods on a benchmark example from the linear bandits literature. For $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^d$, we take $\boldsymbol{x}_1 = \boldsymbol{x}_* = \theta_* = e_1$ and $\boldsymbol{x}_2 = e_2$. The remaining $\boldsymbol{x}_3, \cdots, \boldsymbol{x}_n$ are set so that their first two coordinates

are $\cos(\pi/4(1+\xi))e_1$ and $\sin(\pi/4(1+\xi))e_2$ for $\xi \sim \text{Unif}(-.2, .2)$. We set the threshold $\alpha = 0.5$, $n = 100$, and $d = 25$. Figure B.4a shows that `MELK` outperforms `LSE` when both algorithms are run with their exact confidence widths.

In the implicit setting, this example is especially informative and highlights the importance of designing to choose which arms to sample. Though it is far below α, sampling arm $\boldsymbol{x}_2$ provides the most information about which arms exceed the implicit threshold. Indeed, we see in B.4b that both `MILK` greatly outperforms `LSE-imp` respectively.

B.6 Reducing Experimental Design in an RKHS to a finite dimensional optimization

In this section we describe the use of the kernel trick and Frank-Wolfe to compute the design

$$f(\lambda) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \in C} \|\phi(\boldsymbol{x})\|_{A^\gamma(\lambda)^{-1}}$$

where $C \subset \mathcal{X}$.

Since this is a convex optimization problem on the finite dimensional simplex $\triangle_{\mathcal{X}}$ we employ the Frank-Wolfe algorithm. Note that λ_t is at most t-sparse. The primary challenge is in the computation of the

Algorithm B.2. Frank-Wolfe to minimize f

Input: Arms $\mathcal{X}$, iterations T

$\lambda_0 = \mathbf{e}_1$ (first standard basis vector)
for $\boldsymbol{x} \in \mathcal{A}_t$ **do**
  $x_t \leftarrow \arg\max_{x \in \mathcal{X}} \|\phi(\boldsymbol{x})\|^2_{A^\gamma(\lambda)^{-1}}$
  $g_t = \nabla_{\lambda_{t-1}} \|\phi(\boldsymbol{x}_t)\|^2_{A^\gamma(\lambda)^{-1}}$
  $j_t = \arg\max_{1 \le j \le |\mathcal{X}|} e_i^\top g_t$
  $\eta_t = \frac{1}{t+2}$
  $\lambda_t = (1 - \eta_t)\lambda_{t-1} + \eta_t$
return λ_T

gradient of f. To do so we leverage a small modification of Lemma 1 of (Camilleri et al., 2021a).

Lemma 34. *Assume that λ is s-sparse and (without loss of generality) with its support corresponding to $x_1, \cdots, x_s \in \mathcal{X}$. Then,*

$$\phi(\boldsymbol{x})^\top A^\gamma(\lambda)^{-1} \phi(\boldsymbol{y}) = \frac{k(\boldsymbol{x}, \boldsymbol{y})}{\gamma} - \frac{1}{\gamma} k_\lambda(\boldsymbol{x})^\top (K_\lambda + \gamma I_s)^{-1} k_\lambda(\boldsymbol{y})$$

where $k_\lambda(\cdot) \in \mathbb{R}^s$ with $[k_{\lambda(\boldsymbol{x})}]_i = \sqrt{\lambda_i} k(\boldsymbol{x}_i, \boldsymbol{x})$ for $i \le s$ and $K_\lambda \in \mathbb{R}^{s \times s}$ with $[K_\lambda]_{i,j} = \sqrt{\lambda_i \lambda_j} k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Now, identifying $\mathcal{X}$ with an indexing of its entries, i.e. $\mathcal{X} = \{\boldsymbol{x}^1, \cdots, \boldsymbol{x}^{|\mathcal{X}|}\}$ a computation shows that

$$\mathbf{e}_i^\top [g_t] = -(\phi(\boldsymbol{x}_t) A^\gamma(\lambda_t)^{-1} \phi(\boldsymbol{x}^i))^2$$

which can be computed by the above lemma. Note that computationally, the most difficult step is the inversion of a $t \times t$ matrix at iteration t. For a small number of iterations (<2000), this is not prohibitive.

Appendix C

Appendix for Chap. 4

C.1 Selective Sampling Lower Bound

First, we review the standard argument for best-arm identification lower bounds applied to linear bandits. Fix $\theta_* \in \mathbb{R}^d$ and let $z_* = \arg\max_{z \in \mathcal{Z}} \langle z, \theta_* \rangle$. Define the set $\mathcal{C} = \{\theta \in \mathbb{R}^d : \exists z \in \mathcal{Z} \text{ s.t. } \langle \theta, z - z_* \rangle \geq 0\}$ as those θ in which z_* is note the best arm under θ. We now recall the transportation lemma of (Kaufmann et al., 2016). Under a δ-PAC strategy for finding the best arm for the bandit instance $(\mathcal{X}, \mathcal{Z}, \theta_*)$, let T_x denote the random variable which is the number of times arm x is pulled. In addition let $\mathcal{N}_{\theta,x}$ denote the reward distribution of the arm x of $\mathcal{X}$, i.e. $\mathcal{N}_{\theta,x} = \mathcal{N}(x^\top \theta, 1)$. Then for any δ-PAC algorithm

$$\begin{aligned}
\log(1/2.4\delta) &\leq \min_{\theta \in \mathcal{C}} \sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \mathrm{KL}(\mathcal{N}_{\theta_*,x}, \mathcal{N}_{\theta,x}) \\
&= \min_{\theta \in \mathcal{C}} \sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, \tfrac{1}{2} \|\theta_* - \theta\|^2_{xx^\top} \\
&= \min_{\theta \in \mathcal{C}} \tfrac{1}{2} \|\theta_* - \theta\|^2_{(\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)} \\
&\leq \min_{z \in \mathcal{Z} \setminus z_*} \tfrac{1}{2} \|\theta_* - \theta_z(\epsilon)\|^2_{(\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)}
\end{aligned}$$

where

$$\theta_z(\varepsilon) = \theta_* - \frac{((z_* - z)^\top \theta_* + \varepsilon)(\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)^{-1}(z_* - z)^\top}{(z_* - z)^\top (\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)^{-1}(z_* - z)}$$

for some small ϵ. This is a valid choice since for all $z \in \mathcal{Z} \setminus z_*$ we have $(z_* - z)^\top \theta_z(\varepsilon) = -\varepsilon < 0$ and thus $\theta_z(\varepsilon) \in \mathcal{C}$. A straightforward calculation shows that

$$\|\theta_* - \theta_z(\epsilon)\|^2_{(\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)} = \frac{(\langle z_* - z, \theta_* \rangle + \varepsilon)^2}{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)^{-1}}}$$

so that after rearranging and lettering $\epsilon \to 0$ we have that any δ-PAC algorithm satisfies

$$\max_{z \in \mathcal{Z} \setminus z_*} \frac{2\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \mathbb{E}[T_x] \, xx^\top)^{-1}}}{\langle z_* - z, \theta_* \rangle^2} \log(1/2.4\delta) \leq 1. \tag{C.1}$$

This series of steps will be applied for each bullet point of the theorem.

C.1.1 Proof of Theorem 10, part I

We use the consequence of Lemma 19 of (Kaufmann et al., 2016). Consider a δ-PAC algorithm that sets $P(x) = 1$ for all $x \in \mathcal{X}$ for all time until it exits at time $\mathcal{U}$ after this many unlabelled examples have been observed. If T_x denotes the number of times $x \in \mathcal{X}$ was observed before stopping time $\mathcal{U}$, then by Wald's identity we have that

$$\mathbb{E}[T_x] = \mathbb{E}\left[\sum_{t=1}^{\mathcal{U}} \mathbf{1}\{x_t = x\}\right] = \nu(x)\mathbb{E}[\mathcal{U}].$$

Plugging this into Equation C.1 and rearranging we conclude that

$$\mathbb{E}[\mathcal{U}] \geq \max_{z \in \mathcal{Z} \setminus z_*} \frac{2\|z_* - z\|^2_{(\sum_{x\in\mathcal{X}} \nu(x)\, xx^\top)^{-1}}}{\langle z_* - z, \theta_* \rangle^2} \log(1/2.4\delta) =: \rho(\nu) \log(1/2.4\delta)$$

which concludes the proof of the first bullet.

C.1.2 Proof of Theorem 10, part II

By definition, the (random) number of times we measure x is

$$\mathcal{L}_x = \sum_{s=1}^{\mathcal{U}} \mathbf{1}\{x_s = x, Q_s(x) = 1\}$$

and we want to show that $\mathbb{E}[\mathcal{L}_x] = \nu(x)\mathbb{E}\left[\sum_{\ell=1}^{\mathcal{U}} P_\ell(x)\right]$. To do so, we define

$$M_t = \sum_{s=1}^{t} \left(\mathbf{1}\{x_s = x, Q_s(x) = 1\} - \nu(x)P_s(x)\right)$$

It is easy to check that $P_{t+1} \in \mathcal{F}_t := \{(x_s, y_s, Q_s)\}_{s=1}^t$ and that

$$\mathbb{E}[M_{t+1}|\mathcal{F}_t] = M_t + \mathbb{E}[\mathbf{1}\{x_s = x, Q_s(x) = 1\} - \nu(x)P_s(x)|\mathcal{F}_t] = M_t$$

Applying Doob's equality $\mathbb{E}[M_{\mathcal{U}}] = \mathbb{E}[M_0] = 0$. Consequence:

$$\mathbb{E}[\mathcal{L}_x] = \mathbb{E}\left[\sum_{s=1}^{\mathcal{U}} \mathbf{1}\{x_s = x, Q_s(x) = 1\}\right] = \nu(v)\mathbb{E}\left[\sum_{s=1}^{\mathcal{U}} P_s(x)\right]$$

Define $\alpha(x) := \frac{\mathbb{E}[\sum_{s=1}^{\mathcal{U}} P_s(x)]}{\mathbb{E}[\mathcal{U}]}$ and note that each $\alpha_x \in [0, 1]$. Then $\mathbb{E}[\mathcal{L}_x] = \mathbb{E}[\mathcal{U}]\alpha(x)\nu(x)$ so applying equation (18) of (Kaufmann et al., 2016) again, we have

$$\begin{aligned}
\log(1/2.4\delta) &\leq \min_{\theta \in \mathcal{C}} \sum_{x\in\mathcal{X}} \mathbb{E}[\mathcal{L}_x] \mathrm{KL}(\mathcal{N}_{\theta_*,x}, \mathcal{N}_{\theta,x}) \\
&= \min_{\theta \in \mathcal{C}} \sum_{x\in\mathcal{X}} \mathbb{E}[\mathcal{L}_x] \, \|\theta - \theta_*\|^2_{xx^\top}/2 \\
&= \min_{z \in \mathcal{Z} \setminus z_*} \frac{\langle \theta_*, z_* - z \rangle^2}{2\|z - z_*\|^2_{(\sum_{x\in\mathcal{X}} \mathbb{E}[\mathcal{L}_x] xx^\top)^{-1}}}
\end{aligned}$$

$$= \min_{z \in \mathcal{Z} \setminus z_*} \frac{\langle \theta_*, z_* - z \rangle^2}{2\|z - z_*\|^2_{(\sum_{x \in \mathcal{X}} \nu(x)\alpha(x)xx^\top)^{-1}}} \mathbb{E}[\mathcal{U}].$$

Rearranging, and applying the identity $\mathbb{E}_{X \sim \nu}[\alpha(X)XX^\top] = \sum_{x \in \mathcal{X}} \nu(x)\alpha(x)xx^\top$, the above implies that

$$\mathbb{E}[\mathcal{U}] \geq \max_{z \in \mathcal{Z} \setminus z_*} \frac{2\|z - z_*\|^2_{\mathbb{E}_{X \sim \nu}[\alpha(X)XX^\top]^{-1}}}{\langle \theta_*, z_* - z \rangle^2} \log(1/2.4\delta).$$

Noting that the total expected number of labels is equal to

$$\mathbb{E}[\mathcal{L}] = \sum_{x \in \mathcal{X}} \mathbb{E}[\mathcal{L}_x] = \sum_{x \in \mathcal{X}} \mathbb{E}[\mathcal{U}]\alpha(x)\nu(x) = \mathbb{E}[\mathcal{U}]\, \mathbb{E}_{X \sim \nu}[\alpha(X)]$$

we conclude that

$$\begin{aligned} \mathbb{E}[\mathcal{L}] \geq \min_{\alpha : \mathcal{X} \to [0,1]} \quad & \mathbb{E}[\mathcal{U}]\, \mathbb{E}_{X \sim \nu}[\alpha(X)] \\ \text{subject to} \quad & \mathbb{E}[\mathcal{U}] \geq \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{2\|z - z_*\|^2_{\mathbb{E}_{X \sim \nu}[\alpha(X)XX^\top]^{-1}}}{\langle \theta_*, z_* - z \rangle^2} \log(1/2.4\delta). \end{aligned}$$

The second bullet point result follows by denoting α as P and applying Proposition 4.

C.2 Selective Sampling Algorithm for Known Distribution ν

C.2.1 Proof of Theorem 11, upper bound

At each round ℓ we assume an implementation such that $\widehat{P}_\ell, \widehat{\Sigma}_{\widehat{P}_\ell} \leftarrow$ OPTIMIZEDESIGN$(\mathcal{Z}_\ell, 2^{-\ell}, \tau)$ returns the solution of Equation 4.3 with $\epsilon = 2^{-\ell}$, essentially. More explicitly, let $\epsilon_\ell := 2^{-\ell}$, $B < \infty$ such that $\max_{x \in \mathcal{X}} |\langle x, \theta_* \rangle| \leq B$, and $\sigma < \infty$ such that $\mathbb{E}[(y_s - \langle \theta_*, x_s \rangle)^2 | x_s] \leq \sigma^2$. If

$$\beta_{\delta,\ell} := 16(B^2 + \sigma^2) \log(2\ell^2 |\mathcal{Z}|^2 / \delta)$$

then $\widehat{P}_\ell = P_\ell$ where

$$P_\ell := \text{argmin}_{P : \mathcal{X} \to [0,1]} \mathbb{E}_{X \sim \nu}[P(X)] \text{ subject to } \max_{z, z' \in \mathcal{Z}_\ell} \frac{\|z - z'\|^2_{\mathbb{E}_{X \sim \nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2} \beta_{\delta,\ell} \leq 1$$

and $\widehat{\Sigma}_{\widehat{P}_\ell} := \mathbb{E}_{X \sim \nu}[P_\ell(X)XX^\top]$

We first provide an intermediate lemma on the correctness of Algorithm 4.1 that relies on the feasibility of P_ℓ which we will show shortly.

Lemma 35. *With probability at least $1 - \delta$ we have for all stages $\ell \in \mathbb{N}$ such that P_ℓ is feasible, that $z_* \in \mathcal{Z}_\ell$ and $\max_{z \in \mathcal{Z}_\ell} \langle z_* - z, \theta_* \rangle \leq 4\epsilon_\ell$.*

Proof. Define the event $\mathcal{E}$ as

$$\mathcal{E} := \bigcap_{\ell=1}^{\infty} \bigcap_{z, z' \in \mathcal{Z}_\ell} \left\{ |\langle z - z', \widehat{\theta}_\ell - \theta_* \rangle| \leq \epsilon_\ell \right\}$$

By Lemma 36, we know that $\mathbb{P}(\mathcal{E}) \geq 1-\delta$. Then, the rest of the proof is the same as the one in (Fiez et al., 2019), but we include it here for completeness. Assume that $\mathcal{E}$ holds. Then for any $z' \in \mathcal{Z}_\ell$

$$\begin{aligned}\langle z'-z^*, \widehat{\theta}_\ell\rangle &= \langle z'-z^*, \widehat{\theta}_\ell-\theta^*\rangle + \langle z'-z^*, \theta^*\rangle \\ &= \langle z'-z^*, \widehat{\theta}_\ell-\theta^*\rangle \\ &\leq \epsilon_\ell\end{aligned}$$

so that z^* would survive to round $\mathcal{Z}_{\ell+1}$. And for any $z \in \mathcal{Z}_\ell$ such that $\langle z^*-z, \theta^*\rangle > 2\epsilon_\ell$, we have

$$\begin{aligned}\max_{z'\in\mathcal{Z}_\ell}\langle z'-z, \widehat{\theta}_\ell\rangle &\geq \langle z^*-z, \widehat{\theta}_\ell\rangle \\ &= \langle z^*-z, \widehat{\theta}_\ell-\theta^*\rangle + \langle z^*-z, \theta^*\rangle \\ &> -\epsilon_\ell + 2\epsilon_\ell \\ &= \epsilon_\ell\end{aligned}$$

which implies this z would be kicked out. Note that this implies that $\max_{z\in\mathcal{Z}_{\ell+1}}\langle z^*-z,\theta^*\rangle \leq 2\epsilon_\ell = 4\epsilon_{\ell+1}$. □

We can now prove Theorem 11. After $L := \lceil \log_2(\frac{4}{\Delta})\rceil$ rounds $\mathcal{Z}_\ell = \{z_*\}$ by the above lemma. Thus, the total number of labels requested after L rounds is equal to $\mathcal{L} := \sum_{\ell=1}^{L}\sum_{t=(\ell-1)\tau+1}^{\ell\tau} Q_\ell(x_t)$. By Freedman's inequality (c.f., Theorem 1 of (Beygelzimer et al., 2011)) we have that

$$\sum_{\ell=1}^{L}\sum_{t=(\ell-1)\tau+1}^{\ell\tau} Q_\ell(x_t) \leq 2\sum_{\ell=1}^{L}\tau\mathbb{E}_{X\sim\nu}[P_\ell(X)|\mathcal{Z}_\ell] + \log(1/\delta)$$

We can now bound the expected sample complexity of this algorithm.

$$\begin{aligned}&\sum_{\ell=1}^{L}\tau\mathbb{E}_{X\sim\nu}[P_\ell(X)|\mathcal{Z}_\ell] \\ &= \sum_{\ell=1}^{L}\left[\min_{P:\mathcal{X}\to[0,1]}\tau\mathbb{E}_{X\sim\nu}[P(X)] \quad \text{subject to} \quad \max_{z,z'\in\mathcal{Z}_\ell}\frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2}\beta_{\delta,\ell} \leq 1\right].\end{aligned}$$

Using Lemma 37, we have

$$\begin{aligned}\max_{z,z'\in\mathcal{Z}_\ell}\frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2}\beta_{\delta,\ell} &\leq \beta_{\delta,L}\max_{z,z'\in\mathcal{Z}_\ell}\frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2} \\ &\leq 64\beta_{\delta,L}\max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\langle z-z_*,\theta_*\rangle^2} \\ &=: \max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\langle z-z_*,\theta_*\rangle^2}\beta_\delta\end{aligned}$$

Note that the last line also describes a condition for which a P_ℓ is feasible. Indeed, at round ℓ, a sufficient condition for a feasible P_ℓ (i.e., the RHS $\leq$ 1) is if τ exceeds $\rho(\nu)\beta_\delta$ with $\beta_\delta := 1024(B^2+\sigma^2)\log(2L^2|\mathcal{Z}|^2/\delta)$ and $\rho(\nu) = \max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[XX^\top]^{-1}}}{\langle z-z_*,\theta_*\rangle^2}$, which holds by assumption in the theorem.

Plugging this constraint back into above we have

$$\begin{aligned}
&\sum_{\ell=1}^{L} \tau \mathbb{E}_{X\sim\nu}[P_\ell(X)|\mathcal{Z}_\ell] \\
&\leq \sum_{\ell=1}^{L} \left[\min_{P:\mathcal{X}\to[0,1]} \tau \mathbb{E}_{X\sim\nu}[P(X)] \quad \text{subject to} \quad \max_{z\in\mathcal{Z}\setminus z_*} \frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\langle z-z_*, \theta_*\rangle^2} \beta_\delta \leq 1 \right] \\
&\leq L \min_{\lambda\in\triangle_{\mathcal{X}}} \rho(\lambda)\beta_\delta \quad \text{subject to} \quad \|\lambda/\nu\|_\infty \rho(\lambda)\beta_\delta \leq \tau
\end{aligned}$$

where the last line follows by applying the reparameterization of Proposition 4.

High-probability Events

Lemma 36. *We have* $\mathbb{P}(\mathcal{E}) \geq 1-\delta$.

Proof. For any $\mathcal{V} \subseteq \mathcal{Z}$ and $z, z' \in \mathcal{V}$ define

$$\mathcal{E}_{z,z',\ell}(\mathcal{V}) = \{|\langle z-z', \widehat{\theta}_\ell(\mathcal{V}) - \theta_*\rangle| \leq \epsilon_\ell\}$$

where $\widehat{\theta}_\ell(\mathcal{V})$ is the estimator that would be constructed by the algorithm at stage ℓ with $\mathcal{Z}_\ell = \mathcal{V}$. For fixed $\mathcal{V} \subset \mathcal{Z}$ and $\ell \in \mathbb{N}$ we apply Proposition 3 so that with probability at least $1 - \frac{\delta}{\ell^2|\mathcal{Z}|^2}$ we have that for any $z, z' \in \mathcal{V}$

$$\begin{aligned}
|\langle z-z', \widehat{\theta}_\ell(\mathcal{V}) - \theta_*\rangle| &\leq \|z-z'\|_{\mathbb{E}_{X\sim\nu}[\tau P_\ell(X)XX^\top]^{-1}} \sqrt{16(B^2+\sigma^2)\log(2\ell^2|\mathcal{Z}|^2/\delta)} \\
&\leq \epsilon_\ell
\end{aligned}$$

Noting that $\mathcal{E} := \bigcap_{\ell=1}^\infty \bigcap_{z,z'\in\mathcal{Z}_\ell} \mathcal{E}_{z,z',\ell}(\mathcal{Z}_\ell)$ we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\ell=1}^\infty \bigcup_{z,z'\in\mathcal{Z}_\ell} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{Z}_\ell)\} \right) &\leq \sum_{\ell=1}^\infty \mathbb{P}\left(\bigcup_{z,z'\in\mathcal{Z}_\ell} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{Z}_\ell)\} \right) \\
&= \sum_{\ell=1}^\infty \sum_{\mathcal{V}\subseteq\mathcal{Z}} \mathbb{P}\left(\bigcup_{z,z'\in\mathcal{V}} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{V})\}, \mathcal{Z}_\ell = \mathcal{V} \right) \\
&= \sum_{\ell=1}^\infty \sum_{\mathcal{V}\subseteq\mathcal{Z}} \mathbb{P}\left(\bigcup_{z,z'\in\mathcal{V}} \{\mathcal{E}^c_{z,z',\ell}(\mathcal{V})\} \right) \mathbb{P}(\mathcal{Z}_\ell = \mathcal{V}) \\
&\leq \sum_{\ell=1}^\infty \sum_{\mathcal{V}\subseteq\mathcal{Z}} \tfrac{\delta}{\ell^2|\mathcal{Z}|^2} \binom{|\mathcal{V}|}{2} \mathbb{P}(\mathcal{Z}_\ell = \mathcal{V}) \\
&\leq \sum_{\ell=1}^\infty \sum_{\mathcal{V}\subseteq\mathcal{Z}} \tfrac{\delta}{2\ell^2} \mathbb{P}(\mathcal{Z}_\ell = \mathcal{V}) \leq \delta
\end{aligned}$$

□

C.2.2 Technical Lemmas

The following definition characterizes the RIPS estimator we used in Algorithm 4.1.

Definition 7. *Let $X_1, \dots, X_n$ be i.i.d. random variables with mean $\bar{x}$ and variance ν^2. Let $\delta \in (0,1)$. We say that $\widehat{\mu}(X_1, \dots, X_n)$ is a δ-*robust estimator *if there exist universal constants $c_1, c_0 > 0$ such that if $n \geq c_1 \log(1/\delta)$, then with probability at least $1-\delta$*

$$|\widehat{\mu}(\{X_t\}_{t=1}^n) - \bar{x}| \leq c_0 \sqrt{\frac{\nu^2 \log(1/\delta)}{n}}.$$

Examples of δ-robust estimators include the median-of-means estimator and Catoni's estimator (Lugosi and Mendelson, 2019). This work employs the use of the Catoni estimator which satisfies $|\widehat{\mu}(\{X_t\}_{t=1}^n) - \bar{x}| \leq \sqrt{\frac{2\nu^2 \log(1/\delta)}{n - 2\log(1/\delta)}}$ for $n > 2\log(1/\delta)$ which leads to an optimal leading constant as $n \to \infty$. See (Camilleri et al., 2021a) or (Lugosi and Mendelson, 2019) for more details.

Proposition 3. *Let $x_1, \dots, x_n$ be drawn IID from a distribution ν. Assume that $|\langle \theta, x_s \rangle| \leq B$ and $\mathbb{E}[|\langle \theta, x_s \rangle - y_s|^2] \leq \sigma^2$. Let $P : \mathcal{X} \to [0,1]$ be arbitrary. Let $Q(x_s) \sim$ Bernoulli$(P(x_s))$ independently for all $s \in [n]$. For a given finite set $\mathcal{V} \subset \mathbb{R}^d$ define for any $v \in \mathcal{V}$*

$$w_v = \text{Catoni}(\{\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1} Q(x_s) x_s y_s \rangle\}_{s=1}^n).$$

If $\widehat{\theta} = \arg\min_\theta \max_v \frac{|w_v - \langle \theta, v \rangle|}{\|v\|_{\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}}}$ and $n \geq 4\log(2|\mathcal{V}|/\delta)$, then with probability at least $1-\delta$, it holds that

$$|\langle v, \widehat{\theta} - \theta \rangle| \leq \|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}} \sqrt{16(B^2+\sigma^2)\log(2|\mathcal{V}|/\delta)}$$

Proof. Inspired by (Camilleri et al., 2021a), we note that

$$\begin{aligned}
\max_{v\in\mathcal{V}} \frac{|\langle \widehat{\theta}, v \rangle - \langle \theta, v \rangle|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}} &= \max_{v\in\mathcal{V}} \frac{|\langle \widehat{\theta}, v \rangle - w_v + w_v - \langle \theta, v \rangle|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}} \\
&\leq \max_{v\in\mathcal{V}} \frac{|\langle \widehat{\theta}, v \rangle - w_v|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}} + \max_{v\in\mathcal{V}} \frac{|w_v - \langle \theta, v \rangle|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}} \\
&= \min_\theta \max_{v\in\mathcal{V}} \frac{|\langle \theta, v \rangle - w_v|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}} + \max_{v\in\mathcal{V}} \frac{|w_v - \langle \theta, v \rangle|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}} \\
&\leq 2 \max_{v\in\mathcal{V}} \frac{|\langle \theta, v \rangle - w_v|}{\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}}
\end{aligned}$$

So it suffices to show that each $|\langle \theta, v \rangle - w_v|$ is small. We begin by fixing some $v \in \mathcal{V}$ and bounding the variance of $v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1} Q(x_s) x_s y_s$ for any $s \leq n$ which is necessary to use the robust estimator. For readability purposes, we shorten $\mathbb{E}_{x_s\sim\nu, Q(x_s)\sim P(x_s)}$ as $\mathbb{E}_{x_s,Q}$ in the rest of this proof. Note that

$$\begin{aligned}
&\mathbb{V}\text{ar}_{x_s\sim\nu, Q(x_s)\sim P(x_s)}(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1} Q(x_s) x_s y_s) \\
=&\mathbb{E}_{x_s,Q}[(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1} Q(x_s) x_s y_s)^2] \\
&\quad - \mathbb{E}_{x_s,Q}[v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1} Q(x_s) x_s y_s]^2
\end{aligned}$$

which means we can drop the second term to bound the variance by

$$
\begin{aligned}
&\mathbb{E}_{x_s,Q}[\Big((v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s y_s\Big)^2]\\
&= \mathbb{E}_{x_s,Q}[\Big(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s(x_s^\top\theta+\xi_s)\Big)^2]\\
&= \mathbb{E}_{x_s,Q}[\Big(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s(x_s^\top\theta)\Big)^2]\\
&\qquad + \mathbb{E}_{x_s,Q}[\Big(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s\Big)^2\xi_t^2]\\
&\le B^2\mathbb{E}_{x_s,Q}[\Big(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s\Big)^2]\\
&\qquad + \sigma^2\mathbb{E}_{x_s,Q}[\Big(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s\Big)^2]\\
&= \mathbb{E}_{x_s\sim\nu}\Big[(B^2+\sigma^2)\mathbb{E}_{Q(x_s)\sim P(x_s)}[v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s x_s^\top Q(x_s)\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}v]\Big]\\
&\overset{\text{(i)}}{=} \mathbb{E}_{x_s\sim\nu}\Big[(B^2+\sigma^2)\mathbb{E}_{Q(x_s)\sim P(x_s)}[v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s x_s^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}v]\Big]\\
&\le \mathbb{E}_{x_s\sim\nu}\Big[(B^2+\sigma^2)v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}P(x_s)x_s x_s^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}v]\Big],
\end{aligned}
$$

where we used that $Q(x_s)^2 = Q(x_s)$ in equality (i) above. Thus, we have

$$
\begin{aligned}
&\mathbb{V}\text{ar}(v^\top \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s y_s)\\
\le&(B^2+\sigma^2)v^\top(\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}\mathbb{E}_{x_s\sim\nu}[P(x_s)x_s x_s^\top](\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1})v\\
=&(B^2+\sigma^2)\|v\|^2_{(\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}}
\end{aligned}
$$

By using the property of Catoni estimator stated in Definition 7, we have $c_0 = \sqrt{2}$ and

$$
\begin{aligned}
&|\langle\theta_*, v\rangle - w_v|\\
=&|\text{Catoni}(\{\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s y_s\rangle\}_{s=1}^n) - \mathbb{E}[\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s y_s\rangle]|\\
\le&\sqrt{2}\sqrt{(\mathbb{V}\text{ar}(\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q(x_s)x_s y_s\rangle))\,\frac{\log(\frac{2}{\delta})}{n/2}}\\
&\hfill\text{(with probability at least } 1-\delta \text{ if } n \ge 4\log(2/\delta))\\
\le&\|v\|_{(\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}}\sqrt{\frac{4(B^2+\sigma^2)\log(\frac{2}{\delta})}{n}}\\
=&\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}\sqrt{4(B^2+\sigma^2)\log(2/\delta)}.
\end{aligned}
$$

Finally, the proof is complete by taking union bounding over all $v \in \mathcal{V}$. □

Lemma 37. *Holds*

$$
\max_{z,z'\in\mathcal{Z}_\ell}\frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2} \le 64\max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\langle z-z_*, \theta_*\rangle^2}
$$

Proof. Let $\mathcal{S}_\ell = \{z \in \mathcal{Z} : \langle z_* - z, \theta_*\rangle \le 4\epsilon_\ell\}$. We have

$$
\max_{z,z'\in\mathcal{Z}_\ell}\frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2} \le \max_{z,z'\in\mathcal{S}_\ell}\frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2}
$$

$$
\begin{aligned}
&= 16 \max_{z,z' \in \mathcal{S}_\ell} \frac{\|z - z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X) X X^\top]^{-1}}}{(4\epsilon_\ell)^2} \\
&\le 64 \max_{z \in \mathcal{S}_\ell} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X) X X^\top]^{-1}}}{(4\epsilon_\ell)^2} \\
&= 64 \max_{z \in \mathcal{S}_\ell \setminus z_*} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X) X X^\top]^{-1}}}{\max\{(4\epsilon_\ell)^2, \langle z - z_*, \theta_* \rangle^2\}} \\
&\le 64 \max_{z \in \mathcal{Z} \setminus z_*} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X) X X^\top]^{-1}}}{\langle z - z_*, \theta_* \rangle^2}.
\end{aligned}
$$

□

Reparameterization

Proposition 4. *Fix $\nu \in \triangle_\mathcal{X}$ and any $\lambda \in \triangle_\mathcal{X}$. Define $\|\lambda/\nu\|_\infty = \sup_{x\in\mathcal{X}} \lambda(x)/\nu(x)$ and $\rho(\lambda) = \max_{z\neq z_*} \frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\lambda}[XX^\top]^{-1}}}{\langle z_*-z,\theta_*\rangle^2}$. For any $t, \beta \in \mathbb{R}_+$ the following optimization problems achieve the same value*

- $\min_{P:\mathcal{X}\to[0,1]} t\, \mathbb{E}_{X\sim\nu}[P(X)]$ *subject to* $\max_{z\neq z_*} \frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}}}{\langle z_*-z,\theta_*\rangle^2}\beta \le t$
- $\min_{\lambda\in\triangle_\mathcal{X}} \rho(\lambda)\beta \quad$ *subject to* $\quad \|\lambda/\nu\|_\infty \rho(\lambda)\beta \le t$

Let us first prove a simple lemma.

Lemma 38. *Let $\mathcal{P}$ denote the set of all functions $P : \mathcal{X} \to [0, 1]$. And for any $\nu \in \triangle_\mathcal{X}$ with support $\mathcal{X}$ let $\mathcal{P}' = \{\kappa\lambda_x/\nu_x : \lambda \in \triangle_\mathcal{X}, \kappa \ge 0 : \kappa\lambda_x/\nu_x \in [0, 1]\}$. Then $\mathcal{P} = \mathcal{P}'$.*

Proof. Fix any $P \in \mathcal{P}$. If $\lambda_x = P_x\nu_x/\|P \circ \nu\|_1$ and $\kappa = \|P \circ \nu\|_1$ then $\kappa\lambda/\nu \in \mathcal{P}'$ and is equal to P. This implies $\mathcal{P} \subseteq \mathcal{P}'$.

For the other direction, fix any $\lambda \in \triangle_\mathcal{X}$ and $\kappa \ge 0$ such that $\kappa\lambda_x/\nu_x \in [0, 1]$ for all x. If $P = \kappa\lambda/\nu$ then $P \in \mathcal{P}$ which implies $\mathcal{P}' \subseteq \mathcal{P}$ and concludes the proof. □

Proof of Proposition 4. Using the above lemma we have that

$$
\min_{P:\mathcal{X}\to[0,1]} t\, \mathbb{E}_{X\sim\nu}[P(X)] \quad \text{subject to} \quad \max_{z\neq z_*} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}}}{\langle z_* - z, \theta_* \rangle^2}\beta \le t
$$

is equivalent to

$$
\begin{aligned}
\min_{\kappa\ge0,\lambda\in\triangle_\mathcal{X}} t\, \mathbb{E}_{X\sim\nu}[\kappa\lambda(X)/\nu(X)] \quad \text{subject to} \quad & \max_{z\neq z_*} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[\kappa\lambda(X)/\nu(X)XX^\top]^{-1}}}{\langle z_* - z, \theta_* \rangle^2}\beta \le t \\
& \kappa\lambda(x)/\nu(x) \le 1 \quad \forall x \in \mathcal{X}
\end{aligned}
$$

which is equal to, after simplifying,

$$
\min_{\kappa\ge0,\lambda\in\triangle_\mathcal{X}} t\,\kappa \quad \text{subject to} \quad \max_{z\neq z_*} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\lambda}[XX^\top]^{-1}}}{\langle z_* - z, \theta_* \rangle^2}\beta \le t\kappa
$$

$$\kappa\lambda(x)/\nu(x) \le 1 \quad \forall x \in \mathcal{X}$$

which is equal to

$$\min_{u\ge 0, \lambda\in\triangle_{\mathcal{X}}} u \quad \text{subject to} \quad \rho(\lambda)\beta \le u$$
$$\|\lambda/\nu\|_\infty \le \frac{t}{u}.$$

Note, there exists a feasible (λ, u) precisely when there exists a $\lambda \in \triangle_{\mathcal{X}}$ such that $\|\lambda/\nu\|_\infty \rho(\lambda) \le t$, in which case the optimization problem is equal to

$$\min_{\lambda\in\triangle_{\mathcal{X}}} \rho(\lambda)\beta \quad \text{subject to} \quad \|\lambda/\nu\|_\infty \rho(\lambda)\beta \le t$$

□

C.3 Analysis of the Optimization Problem

C.3.1 Proof of Theorem 13

For simplicity, we will use μ instead of μ_b to denote the number that controls the intensity of barrier function.

The proof relies on analyzing another function $\overline{D} : \mathbb{R}^{d\times d}_{\succeq \mathbf{0}} \mapsto \mathbb{R}$. For simplicity, first, we define

$$h_\Lambda(x) = P_\Lambda(x) - \mu\left(\log(1 - P_\Lambda(x)) + \log(P_\Lambda(x))\right) - P_\Lambda(x) x^\top \Lambda x. \tag{C.2}$$

Recall that our dual objective is $D(\mathbf{\Lambda}) = \mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right] + \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y y$. Since the first term in $\mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right]$ only depends on $\Lambda = \sum_{y\in\mathcal{Y}_\ell}\Lambda_y$, we can consider the following optimization problem.

$$\begin{array}{rl} f(\Lambda) = \max_{\Lambda_y} & \sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y y \\ \text{subject to} & \sum_{y\in\mathcal{Y}_\ell}\Lambda_y = \Lambda \\ & \Lambda_y \succeq \mathbf{0}, \quad \forall y \in \mathcal{Y}_\ell. \end{array} \tag{C.3}$$

Then, the alternative dual objective $\overline{D}(\Lambda)$ is defined as $\overline{D}(\Lambda) = \mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right] + \frac{1}{c_\ell^2} f(\Lambda)$. We can immediately see that maximizing $\overline{D}(\cdot)$ is equivalent to maximizing $D(\cdot)$. In particular, let $\Lambda^* \in \arg\max_{\Lambda\succeq\mathbf{0}} \overline{D}(\Lambda)$ and $\left(\Lambda_y^*\right)_{y\in\mathcal{Y}_\ell}$ be the set of PSD matrices that solve problem (C.3) and evaluate $f(\Lambda^*)$. We can see that $\left(\Lambda_y^*\right)_{y\in\mathcal{Y}_\ell}$ also maximizes $D(\cdot)$. Conversely, for $\mathbf{\Lambda}^* = \left(\Lambda_y^*\right)_{y\in\mathcal{Y}_\ell} \in \arg\max_{\Lambda_y\succeq\mathbf{0},\forall y} D(\mathbf{\Lambda})$, we also have $\sum_{y\in\mathcal{Y}_\ell}\Lambda_y^* \in \arg\max_{\Lambda\succeq\mathbf{0}} \overline{D}(\Lambda)$.

Further, we also define their empirical version D_E and $\overline{D}_E$ with extra i.i.d. samples $x_1, \ldots, x_u$ as

$$D_E(\mathbf{\Lambda}) = \frac{1}{u}\sum_{i=1}^u h_\Lambda(x_i) + \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top\Lambda_y y \quad \text{and} \quad \overline{D}_E(\Lambda) = \frac{1}{u}\sum_{i=1}^u h_\Lambda(x_i) + \frac{1}{c_\ell^2} f(\Lambda). \tag{C.4}$$

Recall that the problem Algorithm 4.2 tries to solve is

$$\begin{array}{rl} \min_P & \mathbb{E}_{X\sim\nu}[P(X) - \mu(\log(1 - P(X)) + \log(P(X)))] \\ \text{subject to} & \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right] \succeq \frac{1}{c_\ell^2} yy^\top, \quad \forall y \in \mathcal{Y}_\ell. \end{array} \tag{C.5}$$

We will restate a more precise version of Theorem 13 and then prove it.

Theorem 23. *Suppose* $\|x\|_2 \le M$ *for any* $x \in \operatorname{supp}(\nu)$ *and* $\Sigma = \mathbb{E}_{X\sim\nu}\left[XX^\top\right]$ *is invertible. Let* $\mathbf{\Lambda}^* \in \arg\max_{\Lambda_y \succeq \mathbf{0}, \forall y} D(\mathbf{\Lambda})$ *and* $\kappa(\Sigma) = \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Sigma)}$ *be condition number. Assume* $\|\Lambda^*\|_F > 0$ *and define* $\omega = \min_{\Gamma\in\mathbb{S}^d:\|\Gamma\|_F=1} \mathbb{E}_{X\sim\nu}\left[\left(X^\top \Gamma X\right)^2\right]$*, where* $\mathbb{S}^d$ *is the set of* $d \times d$ *symmetric matrices. Let* $|\mathcal{Y}_\ell|\, C_\ell^2 = \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell}\|y\|_2^4$.

Then, $\Lambda^* = \sum_{y\in\mathcal{Y}_\ell}\Lambda_y^*$ *is unique. Further, for any* $\epsilon > 0$ *and* $\delta > 0$*, suppose it holds that*

$$\mu \le \min\left\{\sqrt{\frac{3\kappa(\Sigma)\,\|\Lambda^*\|_F\, M^2}{8}\cdot\frac{1+\epsilon}{\epsilon}}, \frac{4}{9}\|\Lambda^*\|_F^2\, M^4, \frac{1}{2\sqrt{3}}\right\}$$

$$K \ge \frac{288\kappa(\Sigma)^2\,|\mathcal{Y}_\ell|^3\,\|\Lambda^*\|_F^4\, M^4(M^4+C_\ell^2)\cdot\left(2\,\|\Lambda^*\|_F\, M^2+1\right)^4\log(6/\delta)}{\omega^2\mu^6}\cdot\left(\frac{1+\epsilon}{\epsilon}\right)^2$$

$$u \ge \frac{576\kappa(\Sigma)^2\,\|\Lambda^*\|_F^2\, M^8\cdot\left(2\,\|\Lambda^*\|_F\, M^2+1\right)^4\log(6/\delta)}{\omega^2\mu^6}\cdot\left(\frac{1+\epsilon}{\epsilon}\right)^2.$$

Then, with probability at least $1-\delta$*, Algorithm 4.2 will output* $\widetilde{\Lambda}$ *that satisfies*

- $y^\top\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)XX^\top\right]^{-1}y \le (1+\epsilon)c_\ell^2, \quad \forall y\in\mathcal{Y}_\ell.$
- $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right] \le \mathbb{E}_{X\sim\nu}\left[\widetilde{P}(X)\right] + 4\sqrt{\mu}$*, where* $\widetilde{P}$ *is the optimal solution to problem* (C.13)*.*

Proof. **First Bullet Point.** Fix some $\epsilon > 0$. Let $\hat{\mathbf{\Lambda}}$ and corresponding $\hat{\Lambda} = \sum_{y\in\mathcal{Y}_\ell}\hat{\Lambda}_y$ be the parameters obtained by Algorithm 4.2 just before the re-scaling step, which means that at line 9 of Algorithm 4.2, the assignment of $\hat{\Lambda}_y$ to each $y \in \mathcal{Y}_\ell$ has been optimized by solving problem (C.3). That is, we have $D(\hat{\mathbf{\Lambda}}) = \overline{D}(\hat{\Lambda})$ and $D_E(\hat{\mathbf{\Lambda}}) = \overline{D}_E(\hat{\Lambda})$. Let $\widetilde{\mathbf{\Lambda}}$ and $\widetilde{\Lambda}$ be the ones after the re-scaling step. Then, by Theorem 3.13 of (Orabona, 2019), with probability at least $1-\frac{\delta}{3}$, it holds that

$$\overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) = D(\mathbf{\Lambda}^*) - D(\hat{\mathbf{\Lambda}}) \le \frac{\mathrm{Reg}(K) + 2\sqrt{2K\log(6/\delta)}}{K},$$

where $\mathrm{Reg}(K)$ is the regret of running projected stochastic gradient ascent for K steps with η_k specified in Algorithm 4.2. Meanwhile, by Theorem 4.14 of (Orabona, 2019) also, we have $\mathrm{Reg}(K) = \sqrt{2}B^2\sqrt{\sum_{k=1}^K\sum_{y\in\mathcal{Y}_\ell}\|g_{k,y}\|_2^2}$, where $B = \sqrt{|\mathcal{Y}_\ell|}\,\|\Lambda^*\|_F$ bound the norm of $\mathbf{\Lambda}^* = \left(\Lambda_y^*\right)_{y\in\mathcal{Y}_\ell}$. Since $g_{k,y} = \frac{yy^\top}{c_\ell^2} - P_{\hat{\Lambda}^{(k)}}(x_k)x_kx_k^\top$, we can easily get $\sum_{y\in\mathcal{Y}_\ell}\|g_{k,y}\|_2^2 \le 2\,|\mathcal{Y}_\ell|\,M^4 + \frac{2}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell}\|y\|_2^4 = 2\,|\mathcal{Y}_\ell|\,M^4 + 2\,|\mathcal{Y}_\ell|\,C_\ell^2$. Thus, we have

$$\mathrm{Reg}(K) \le 2\,|\mathcal{Y}_\ell|\,\|\Lambda^*\|_F^2\sqrt{|\mathcal{Y}_\ell|\,M^4 + |\mathcal{Y}_\ell|\,C_\ell^2}\cdot\sqrt{K} := C_{\mathrm{Reg}}\sqrt{K} \tag{C.6}$$

$$\implies \overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) \le \frac{C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}}{\sqrt{K}}, \tag{C.7}$$

We now consider the effect of using u i.i.d. samples in the re-scaling step. First, since re-scaling always increases the function value, we must have $D_E(\hat{\mathbf{\Lambda}}) \le D_E(\widetilde{\mathbf{\Lambda}})$. Meanwhile, since $D_E(\hat{\mathbf{\Lambda}}) = \overline{D}_E(\hat{\Lambda})$, by Lemma 44, we have $D_E(\hat{\mathbf{\Lambda}}) = \overline{D}_E(\hat{\Lambda})$, which together implies $\overline{D}_E(\hat{\Lambda}) \le \overline{D}_E(\widetilde{\Lambda})$.

By Lemma 39, we know that Λ^* is unique and as long as $\mu \le \frac{1}{2\sqrt{3}}$, $\overline{D}(\Lambda)$ is G-strongly concave with respect to ℓ_2 norm over $\mathcal{S} = \{\Lambda \succeq \mathbf{0} : \|\Lambda\|_F \le 2\,\|\Lambda^*\|_F\}$, where G is defined in Eq. (C.14). Thus, by

Lemma 45, if K is large enough such that

$$\overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) \leq \frac{C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}}{\sqrt{K}} \leq \frac{G\left\|\Lambda^*\right\|_F}{2},$$

then $\left\|\hat{\Lambda} - \Lambda^*\right\|_F \leq \left\|\Lambda^*\right\|_F$, which implies $\left\|\hat{\Lambda}\right\|_F \leq 2\left\|\Lambda^*\right\|_F$. That is, $\hat{\Lambda} \in \mathcal{S}$. Then, under this condition, by using Lemma 42, when $\mu \leq \frac{4}{9}\left\|\Lambda^*\right\|_F M^4$ and

$$u \geq \left(\frac{6\kappa(\Sigma)\left\|\Lambda^*\right\|_F M^4\left(2+\sqrt{2\log(6/\delta)}\right)}{G\mu^2}\cdot\frac{1+\epsilon}{\epsilon}\right)^2, \tag{C.8}$$

for $\widetilde{\Lambda}$ after re-scaling, with probability at least $1-\frac{\delta}{3}$, it holds simultaneously that

$$\left|\overline{D}_E(\hat{\Lambda}) - \overline{D}(\hat{\Lambda})\right| \leq \frac{G\mu^2}{3M^2\kappa(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon} \quad \text{and} \quad \left|\overline{D}_E(\widetilde{\Lambda}) - \overline{D}(\widetilde{\Lambda})\right| \leq \frac{G\mu^2}{3M^2\kappa(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon} \tag{C.9}$$

$$\begin{aligned}
\implies \overline{D}(\Lambda^*) - \overline{D}(\widetilde{\Lambda}) &\leq \overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) + \overline{D}(\hat{\Lambda}) - \overline{D}(\widetilde{\Lambda}) \\
&\leq \overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) + \overline{D}(\hat{\Lambda}) - \overline{D}_E(\hat{\Lambda}) + \overline{D}_E(\widetilde{\Lambda}) - \overline{D}(\widetilde{\Lambda}) \\
&\qquad \text{(Since } \overline{D}_E(\hat{\Lambda}) \leq \overline{D}_E(\widetilde{\Lambda})\text{)} \\
&\leq \frac{C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}}{\sqrt{K}} + \frac{2G\mu^2}{3M^2\kappa(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon}. \qquad \text{(By Eq. (C.7) and (C.9))}
\end{aligned}$$

Since $\widetilde{\Lambda}$ is a smaller re-scaling of $\hat{\Lambda}$, we have $\widetilde{\Lambda} \in \mathcal{S}$, which implies $\frac{G}{2}\left\|\Lambda^* - \widetilde{\Lambda}\right\|_F \leq \overline{D}(\Lambda^*) - \overline{D}(\widetilde{\Lambda})$ by property of strongly concave function (Bertsekas, 2009). Therefore, by Lemma 46, to guarantee an at most ϵ multiplicative constraint violation, it is sufficient to choose K such that

$$\begin{aligned}
\frac{G}{2}\left\|\Lambda^* - \widetilde{\Lambda}\right\|_F &\leq \overline{D}(\Lambda^*) - \overline{D}(\widetilde{\Lambda}) \\
&\leq \frac{C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}}{\sqrt{K}} + \frac{2G\mu^2}{3M^2\kappa(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon} \\
&\leq \min\left\{\frac{4G\mu^2}{3M^2\kappa(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon}, \frac{G\left\|\Lambda^*\right\|_F}{2}\right\} \\
&= \frac{4G\mu^2}{3M^2\kappa(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon}. \qquad \text{(If } \mu \leq \sqrt{\tfrac{3\kappa(\Sigma)\|\Lambda^*\|_F M^2}{8}\cdot\tfrac{1+\epsilon}{\epsilon}}\text{)}
\end{aligned}$$

An algebraic rearrangement gives us

$$K \geq \left(\frac{3\kappa(\Sigma)M^2\left(C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}\right)}{2G\mu^2}\cdot\frac{1+\epsilon}{\epsilon}\right)^2. \tag{C.10}$$

Second Bullet Point. We then prove the upper bound for primal objective value $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right]$, which explains the reason why an extra re-scaling step is needed. Define $g(s) = D_E(s\cdot\widetilde{\boldsymbol{\Lambda}})$. By construction, we

know that $g(s)$ is maximized at $s=1$ because $\widetilde{\mathbf{\Lambda}} = s^* \cdot \hat{\mathbf{\Lambda}}$, where $s^* = \arg\max_{s\in[0,1]} D_E(s\cdot\hat{\mathbf{\Lambda}})$. Therefore, we have $g'(1) \geq 0$, which in turn gives us

$$g'(1) = \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top \widetilde{\Lambda}_y y - \frac{1}{u}\sum_{i=1}^{u} P_{\widetilde{\Lambda}}(x_i) x_i^\top \widetilde{\Lambda} x_i \geq 0.$$

By the concentration inequality in Lemma 42, we know that when

$$u \geq \left(\frac{2\left\|\Lambda^*\right\|_F M^2\left(\left\|\Lambda^*\right\|_F M^2 + \mu\sqrt{2\log(6/\delta)}\right)}{\mu^{3/2}}\right)^2, \tag{C.11}$$

with probability at least $1-\frac{\delta}{3}$, it holds that

$$\begin{aligned}
&\left|\mathbb{E}_{X\sim\nu}\left[P_\Lambda(X)X^\top\Lambda X\right] - \frac{1}{u}\sum_{i=1}^{u} P_\Lambda(x_i)x_i^\top\Lambda x_i\right| \leq \sqrt{\mu} \\
\implies &\frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top\widetilde{\Lambda}_y y - \mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)X^\top\widetilde{\Lambda}X\right] + \sqrt{\mu} \geq 0.
\end{aligned} \tag{C.12}$$

Now, let $\widetilde{P}$ be the optimal solution of problem (C.13) and $\hat{P}$ be the optimal solution of the same problem with bound constraint $\mu \leq P(x) \leq 1-\mu$.

$$\begin{aligned}
\min_P \quad & \mathbb{E}_{X\sim\nu}\left[P(X)\right] \\
\text{subject to} \quad & y^\top \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right]^{-1} y \leq c_\ell^2, \quad \forall y\in\mathcal{Y}_\ell, \\
& 0 \leq P(x) \leq 1-\mu, \quad \forall x\in\mathcal{X}.
\end{aligned} \tag{C.13}$$

Then, we can notice that

$$\begin{aligned}
&\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right] \\
\leq &\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X) - \mu(\log(1-P_{\widetilde{\Lambda}}(X)) + \log(P_{\widetilde{\Lambda}}(X)))\right] \\
\leq &\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X) - \mu(\log(1-P_{\widetilde{\Lambda}}(X)) + \log(P_{\widetilde{\Lambda}}(X)))\right] \\
&\quad + \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top\widetilde{\Lambda}_y y - \mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)X^\top\widetilde{\Lambda}X\right] + \sqrt{\mu} && \text{(By Eq. (C.12))} \\
= &\inf_P \mathcal{L}(P, \widetilde{\mathbf{\Lambda}}) + \sqrt{\mu} && \text{(By definition of Lagrangian function and how we solve for } P_\Lambda) \\
\leq &\max_{\Lambda_y\succeq\mathbf{0},\forall y\in\mathcal{Y}_\ell} \inf_P \mathcal{L}\left(P, \mathbf{\Lambda}\right) + \sqrt{\mu} \\
= &\mathbb{E}_{X\sim\nu}\left[P_{\Lambda^*}(X) - \mu(\log(1-P_{\Lambda^*}(X)) + \log(P_{\Lambda^*}(X)))\right] + \sqrt{\mu} \\
\leq &\mathbb{E}_{X\sim\nu}\left[\hat{P}(X) - \mu\log(1-\hat{P}(X))\right] - \mu\log\left(\hat{P}(X)\right) + \sqrt{\mu} && \text{(Since } \hat{P} \text{ is feasible to problem (C.5))} \\
\leq &\mathbb{E}_{X\sim\nu}\left[\hat{P}(X)\right] + 3\sqrt{\mu}, && \text{(Since } -a\log(a) \leq \sqrt{a} \text{ for } a\in(0,1)) \\
\leq &\mathbb{E}_{X\sim\nu}\left[\widetilde{P}(X)\right] + 4\sqrt{\mu}. && \text{(Since } \hat{P}(x) \text{ can have at most } \mu \text{ more contribution than } \widetilde{P})
\end{aligned}$$

Therefore, in summary, Suppose K and u satisfy conditions specified in Eq. (C.10), (C.8) and (C.11) and $\mu \leq \min\left\{\sqrt{\frac{3\kappa(\Sigma)\|\Lambda^*\|_F M^2}{8}\cdot\frac{1+\epsilon}{\epsilon}}, \frac{4}{9}\left\|\Lambda^*\right\|_F^2 M^4, \frac{1}{2\sqrt{3}}\right\}$, where C_{Reg} and G are defined in Eq. (C.6) and

(C.14), respectively. Then. by applying a simple union bound, with probability at least $1-\delta$, the output of Algorithm 4.2 $\widetilde{\Lambda}$ satisfies $y^\top \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right]^{-1} y \le (1+\epsilon)c_\ell^2, \forall y \in \mathcal{Y}_\ell$ and $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right] \le \mathbb{E}_{X\sim\nu}\left[\widetilde{P}(X)\right] + 4\sqrt{\mu}$. □

C.3.2 Relevant Lemmas

Strong Concavity of $\overline{D}(\Lambda)$

Lemma 39. *As long as $\mu \le \frac{1}{2\sqrt{3}}$, $\overline{D}(\Lambda)$ is G-strongly concave with respect to ℓ_2-norm on the bounded region $\mathcal{S} = \{\Lambda \succeq \mathbf{0} : \|\Lambda\|_F \le 2\|\Lambda^*\|_F\}$ with coefficient*

$$G = \frac{\mu}{2\left(2\|\Lambda^*\|_F M^2 + 1\right)^2} \cdot \min_{\Gamma\in\mathbb{S}^d:\|\Gamma\|_F=1} \mathbb{E}_{X\sim\nu}\left[\left(X^\top \Gamma X\right)^2\right]. \tag{C.14}$$

Because of this, as a corollary, Λ^* will be unique.

Proof. By Lemma 40, since $f(\Lambda)$ is concave in Λ, it is sufficient to prove that $\mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right]$ is G-strongly concave on $\mathcal{S}$, where $h_\Lambda(x)$ is defined in Eq. (C.2). Then, we have

$$-\nabla^2_\Lambda \mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right] = \mathbb{E}_{X\sim\nu}\left[\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}(X)\mathrm{vec}\left(XX^\top\right)\mathrm{vec}\left(XX^\top\right)^\top\right].$$

Since $\|x\|_2 \le M$, for any $\Lambda \in \mathcal{S}$, we have $q_\Lambda(x) = x^\top \Lambda x - 1 \le 2\|\Lambda^*\|_F M^2 + 1$. By Lemma, 48, we know that if $12\mu^2 \le \left(2\|\Lambda^*\|_F M^2 + 1\right)^2$, which can be done by choosing $\mu \le \frac{1}{2\sqrt{3}}$, we have $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}(x) \ge \frac{\mu}{2\left(2\|\Lambda^*\|_F M^2+1\right)^2}$ for any $x \in \mathcal{X}$ and $\Lambda \in \mathcal{S}$. Therefore, we have

$$-\nabla^2_\Lambda \mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right] \succeq \gamma \cdot \mathbb{E}_{X\sim\nu}\left[\mathrm{vec}\left(XX^\top\right)\mathrm{vec}\left(XX^\top\right)^\top\right]$$

Now, let $\mathbb{S}$ be the set of all $d\times d$ symmetric matrices. It is obvious that $\mathbb{S}$ is a subspace of the vector space of all $d\times d$ matrices and $\mathcal{S} \subseteq \mathbb{S}$. Thus, by applying Lemma 41, we can conclude that $\mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right]$ is G-strongly concave on $\mathcal{S}$ with respect to ℓ_2 norm and

$$\begin{aligned} G &= \frac{\mu}{2\left(2\|\Lambda^*\|_F M^2 + 1\right)^2} \cdot \min_{\Gamma\in\mathbb{S}^d:\|\Gamma\|_F=1} \mathrm{vec}(\Gamma)^\top \mathbb{E}_{X\sim\nu}\left[\mathrm{vec}\left(XX^\top\right)\mathrm{vec}\left(XX^\top\right)^\top\right]\mathrm{vec}(\Gamma) \\ &= \frac{\mu}{2\left(2\|\Lambda^*\|_F M^2 + 1\right)^2} \cdot \min_{\Gamma\in\mathbb{S}^d:\|\Gamma\|_F=1} \mathbb{E}_{X\sim\nu}\left[\left(X^\top\Gamma X\right)^2\right]. \end{aligned}$$

Thus the proof is complete. □

Lemma 40. *$f(\Lambda)$ defined in Eq.* (C.3) *is concave in Λ.*

Proof. To show its concavity, consider $\Lambda^{(1)} \succeq \mathbf{0}$, $\Lambda^{(2)} \succeq \mathbf{0}$ and some $\gamma \in (0,1)$. Let $(\Lambda_y^{(i)})_{y\in\mathcal{Y}_\ell}$ be the optimal solution obtained by evaluating $f(\Lambda^{(i)})$ for $i \in \{1,2\}$. Then, we can notice that

$$\gamma f(\Lambda^{(1)}) + (1-\gamma) f(\Lambda^{(2)}) = \gamma \sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y^{(1)} y + (1-\gamma)\sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y^{(2)} y$$

$$= \sum_{y \in \mathcal{Y}_\ell} y^\top (\gamma \Lambda_y^{(1)} + (1-\gamma)\Lambda_y^{(2)}) y$$
$$\leq f(\gamma \Lambda^{(1)} + (1-\gamma)\Lambda^{(2)}).$$

The last inequality above holds because $\sum_{y \in \mathcal{Y}_\ell} \Lambda_y^{(i)} = \Lambda^{(i)}$ for $i \in \{1,2\}$ and thus $\sum_{y \in \mathcal{Y}_\ell} \left(\gamma \Lambda_y^{(1)} + (1-\gamma)\Lambda_y^{(2)}\right) = \gamma \Lambda^{(1)} + (1-\gamma)\Lambda^{(2)}$, which means that $(\gamma \Lambda_y^{(1)} + (1-\gamma)\Lambda_y^{(2)})_{y \in \mathcal{Y}_\ell}$ is a feasible solution for problem (C.3) with parameter $\gamma \Lambda^{(1)} + (1-\gamma)\Lambda^{(2)}$. Therefore, we can conclude that $f(\Lambda)$ is concave in Λ. □

Lemma 41. *Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a convex and twice differentiable function in $\mathbb{R}^d$. If for some subspace $S \subseteq \mathbb{R}^d$, we have $\min_{w \in S : \|w\|_2 = 1} w^\top \nabla^2 f(x) w \geq \sigma > 0$, $\forall x \in S$, then f is σ-strongly convex with respect to ℓ_2-norm on S.*

Proof. Suppose S has dimension m and let $v_1, \ldots, v_m$ be a set of orthonormal basis that span S. Then, for each $x \in S$, there exists unique $z \in \mathbb{R}^m$ such that $x = Vz$, where $V = \begin{bmatrix} v_1 & \ldots & v_m \end{bmatrix}$. That is, there is one-to-one correspondence between S and $\mathbb{R}^m$.

Now, we define $g : \mathbb{R}^m \mapsto \mathbb{R}$ as $g(z) = f(Vz)$. It is easy to compute $\nabla^2 g(z) = V^\top \nabla^2 f(Vz) V$. Then, notice that for any $w' \in \mathbb{R}^m$ such that $\|w'\|_2 = 1$, we have $Vw' \in S$ and $\|Vw'\|_2 = \sqrt{w'^\top V^\top V w'} = \sqrt{w'^\top w'} = 1$. Thus, we have

$$\min_{w' \in \mathbb{R}^m : \|w'\|_2 = 1} w'^\top \nabla^2 g(z) w' = \min_{w' \in \mathbb{R}^m : \|w'\|_2 = 1} w'^\top V^\top \nabla^2 f(Vz) V w'$$
$$= \min_{w \in S : \|w\|_2 = 1} w^\top \nabla^2 f(Vz) w \geq \sigma.$$

Therefore, g is σ-strongly convex with respect to ℓ_2 norm. Then, for any $x_1, x_2 \in S$, there exists unique $z_1, z_2 \in \mathbb{R}^m$ such that $x_1 = Vz_1$ and $x_2 = Vz_2$. Notice that $\|z_1 - z_2\|_2 = \|x_1 - x_2\|_2$ since V preserves the norm. Further, by definition of strong convexity, for any $\alpha \in [0,1]$, we have

$$g(\alpha z_1 + (1-\alpha) z_2) + \frac{\sigma}{2} \alpha (1-\alpha) \|z_1 - z_2\|_2^2 \leq \alpha g(z_1) + (1-\alpha) g(z_2)$$
$$\implies f(\alpha V z_1 + (1-\alpha) V z_2) + \frac{\sigma}{2} \alpha (1-\alpha) \|x_1 - x_2\|_2^2 \leq \alpha f(V z_1) + (1-\alpha) f(V z_2)$$
$$\implies f(\alpha x_1 + (1-\alpha) x_2) + \frac{\sigma}{2} \alpha (1-\alpha) \|x_1 - x_2\|_2^2 \leq \alpha f(x_1) + (1-\alpha) f(x_2).$$

Thus, f is also σ-strongly convex with respect to ℓ_2 norm on S. □

Concentration Inequalities

Lemma 42. *Let $x_1, \ldots, x_u \sim \nu$ be i.i.d. samples. If $\left\|\hat{\Lambda}\right\|_F \leq 2 \|\Lambda^*\|_F$, $\|x\|_2 \leq M$ for any $x \in \mathcal{X}$ and $\mu \leq \frac{4}{9} \|\Lambda^*\|_F^2 M^4$, then with probability at least $1 - \frac{2\delta}{3}$, it holds for any $\Lambda \in \Theta = \left\{ s \cdot \hat{\Lambda} : s \in [0,1] \right\}$ simultaneously that*

$$\left| \mathbb{E}_{X \sim \nu} \left[h_\Lambda(X) \right] - \frac{1}{u} \sum_{i=1}^u h_\Lambda(x_i) \right| \leq \frac{2 \|\Lambda^*\|_F M^2 \left(2 + \sqrt{2 \log(6/\delta)}\right)}{\sqrt{u}}$$
$$\left| \mathbb{E}_{X \sim \nu} \left[P_\Lambda(X) X^\top \Lambda X \right] - \frac{1}{u} \sum_{i=1}^u P_\Lambda(x_i) x_i^\top \Lambda x_i \right| \leq \frac{2 \|\Lambda^*\|_F M^2 \left(\|\Lambda^*\|_F M^2 + \mu \sqrt{2 \log(6/\delta)} \right)}{\mu \sqrt{u}}.$$

Proof. To prove the first inequality, first, notice that we have $h_\Lambda(x) = -P_\Lambda(x)q_\Lambda(x) - \mu\left(\log(1-P_\Lambda(x)) + \log(P_\Lambda(x))\right)$, where $q_\Lambda(x) = x^\top \Lambda x - 1$. Since $P_\Lambda(x)$, defined in Eq. (4.7), explicitly only depends on $q_\Lambda(x)$ instead of x directly, we can treat h_Λ as a function of q_Λ and define a function class $\mathcal{F} = \left\{x \mapsto x^\top (s \cdot \hat{\Lambda}) x : s \in [0,1]\right\}$. It is well-known that if h_Λ is L_1-Lipschitz in q_Λ and $|h_\Lambda(x)| \le R_1$ for any $\Lambda \in \Theta$ and $x \sim \nu$, then, with probability at least $1 - \frac{\delta}{3}$, it holds simultaneously for all $\Lambda \in \Theta$ that (Bartlett and Mendelson, 2002; Mohri et al., 2018)

$$\left| \mathbb{E}_{X\sim\nu}\left[h_\Lambda(X)\right] - \frac{1}{u}\sum_{i=1}^{u} h_\Lambda(x_i) \right| \le 2L_1 \cdot \mathcal{R}_u(\mathcal{F}) + R_1\sqrt{\frac{2\log(6/\delta)}{u}}, \tag{C.15}$$

where $\mathcal{R}_u(\mathcal{F})$ is the Rademacher complexity of $\mathcal{F}$.

To find L_1, we can compute

$$\begin{aligned}
\frac{\mathrm{d}h_\Lambda}{\mathrm{d}q_\Lambda} &= -\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda} q_\Lambda - P_\Lambda + \frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}\left(\frac{\mu}{1-P_\Lambda} - \frac{\mu}{P_\Lambda}\right) \\
&= -\frac{\mathrm{d}P_\Lambda}{\mathrm{d}\cdot q_\Lambda} q_\Lambda - P_\Lambda + \frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}\cdot q_\Lambda \qquad \text{(Since } P_\Lambda \text{ satisfies Eq. (4.6))} \\
&= -P_\Lambda
\end{aligned}$$

Therefore, we have $\frac{\mathrm{d}h_\Lambda}{\mathrm{d}q_\Lambda} \in \left[-1, -\frac{\mu}{3}\right]$ by Lemma 48. Therefore, we can set $L_1 = 1$.

Let h_0 be the value of h_Λ when $q_\Lambda = -1$, which means $x^\top \Lambda x = 0$. To find R_1, notice that since $\frac{\mathrm{d}h_\Lambda}{\mathrm{d}q_\Lambda} \in \left[-1, -\frac{\mu}{3}\right]$, we must have $-q_\Lambda + h_0 \le h_\Lambda \le -\frac{\mu}{3}q_\Lambda + h_0$. By Lemma 48, we know that $h_0 \in \left[0, 2\sqrt{\mu}\right]$. Therefore, we have $-x^\top \Lambda x \le h_\Lambda(x) \le -\frac{\mu}{3}x^\top \Lambda x + 3\sqrt{\mu}$ for any $x \in \mathcal{X}$ and $\Lambda \in \Theta$. Since $\|\Lambda\|_F \le \left\|\hat{\Lambda}\right\|_F \le 2\left\|\Lambda^*\right\|_F$, we have $|h_\Lambda(x)| \le 2\left\|\Lambda^*\right\|_F M^2 := R_1$, which holds when $\mu \le \frac{4}{9}\left\|\Lambda^*\right\|_F^2 M^4$. Then, by Lemma 43, we know that $\mathcal{R}_u(\mathcal{F}) \le \frac{2\|\Lambda^*\|_F M^2}{\sqrt{u}}$. Thus, plugging in values of L_1, R_1 and $\mathcal{R}_u(\mathcal{F})$ into Eq. (C.15) gives our first concentration inequality.

We can basically follow exactly the same strategy to prove the second concentration inequality. In particular, define $\tilde{h}_\Lambda(x) = P_\Lambda(x)x^\top \Lambda x = P_\Lambda(x)q_\Lambda(x) + P_\Lambda(x)$. Then, with probability at least $1 - \frac{\delta}{3}$, it holds simultaneously for any $\Lambda \in \Theta$ that

$$\left| \mathbb{E}_{X\sim\nu}\left[\tilde{h}_\Lambda(X)\right] - \frac{1}{u}\sum_{i=1}^{u} \tilde{h}_\Lambda(x_i) \right| \le 2L_2 \cdot \mathcal{R}_u(\mathcal{F}) + R_2\sqrt{\frac{2\log(6/\delta)}{u}}, \tag{C.16}$$

where $\left|\tilde{h}_\Lambda(x)\right| \le R_2$ for any $x \in \mathcal{X}$, $\Lambda \in \Theta$ and $\tilde{h}_\Lambda$ is L_2-Lipschitz in q_Λ.

To find L_2, we can compute

$$\frac{\mathrm{d}\tilde{h}_\Lambda}{\mathrm{d}q_\Lambda} = P_\Lambda + \frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}\cdot x^\top \Lambda x.$$

By Lemma 48, we know that $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda} \in \left[0, \frac{1}{8\mu}\right]$. Thus, we have $\left|\frac{\mathrm{d}\tilde{h}_\Lambda}{\mathrm{d}q_\Lambda}\right| \le 1 + \frac{\|\Lambda^*\|_F M^2}{4\mu} := L_2$. It is obvious that $\tilde{h}_\Lambda(x) \le 2\left\|\Lambda^*\right\|_F M^2 := R_2$. Thus, by plugging the values of L_2, R_2 and $\mathcal{R}_u(\mathcal{F})$ into Eq. (C.16), we can obtain the second concentration inequality.

Finally, both concentration inequalities hold simultaneously with probability at least $1 - \frac{2\delta}{3}$ by a simple union bound. □

Lemma 43. *If* $\left\|\hat{\Lambda}\right\|_F \le 2\left\|\Lambda^*\right\|_F$*, then, we have* $\mathcal{R}_u(\mathcal{F}) \le \sqrt{\frac{\mathbb{E}_{X\sim\nu}\left[(X^\top \hat{\Lambda} X)^2\right]}{u}} \le \frac{2\|\Lambda^*\|_F M^2}{\sqrt{u}}$*, where* $\mathcal{F} = \left\{x \mapsto x^\top (s \cdot \hat{\Lambda}) x : s \in [0,1]\right\}$.

Proof. Let $\sigma_1, \ldots, \sigma_u$ be i.i.d. Rademacher random variables, which are uniform over $\{-1, +1\}$. Let $x_1, \ldots, x_u \sim \nu$ be i.i.d. samples. Then, by definition of Rademacher complexity, we have

$$\begin{aligned}
\mathcal{R}_u(\mathcal{F}) &= \mathbb{E}\left[\sup_{q\in\mathcal{F}} \frac{1}{u}\sum_{i=1}^{u}\sigma_i q(x_i)\right] \\
&= \mathbb{E}\left[\sup_{s\in[0,1]} \frac{1}{u}\sum_{i=1}^{u}\sigma_i x_i^\top (s\hat{\Lambda})x_i\right] && \text{(By definition of } \mathcal{F}) \\
&\overset{\text{(i)}}{=} \frac{1}{u}\mathbb{E}\left[\mathbf{1}\left\{\sum_{i=1}^{n}\sigma_i x_i^\top \hat{\Lambda} x_i \geq 0\right\}\sum_{i=1}^{n}\sigma_i x_i^\top \hat{\Lambda} x_i\right]. \\
&\leq \frac{1}{u}\mathbb{E}\left[\left|\sum_{i=1}^{u}\sigma_i x_i^\top \hat{\Lambda} x_i\right|\right] \\
&\leq \frac{1}{u}\sqrt{\mathbb{E}\left[\left(\sum_{i=1}^{u}\sigma_i x_i^\top \hat{\Lambda} x_i\right)^2\right]} && \text{(By Jensen's inequality)} \\
&= \frac{1}{u}\sqrt{\mathbb{E}\left[\sum_{i=1}^{u}\left(x_i^\top \hat{\Lambda} x_i\right)^2\right]} && \text{(Since } \sigma_i\text{'s are i.i.d. and } \mathbb{E}[\sigma_i] = 0) \\
&= \sqrt{\frac{\mathbb{E}_{X\sim\nu}\left[\left(X^\top \hat{\Lambda} X\right)^2\right]}{u}} \leq \frac{2\left\|\Lambda^*\right\|_F M^2}{\sqrt{u}}.
\end{aligned}$$

Here, the equality (i) holds because when $\sum_{i=1}^{n}\sigma_i x_i^\top \hat{\Lambda} x_i < 0$, the supremum over $s \in [0,1]$ will be obtained by taking $s = 0$; otherwise, it will be obtained by taking $s = 1$. □

Other Lemmas

The following lemma basically shows that $f(\Lambda)$ is linear in scalar multiplication.

Lemma 44. *If $D_E(\hat{\mathbf{\Lambda}}) = \overline{D}_E(\hat{\Lambda})$, with $\hat{\Lambda} = \sum_{y\in\mathcal{Y}_\ell}\hat{\Lambda}_y$, then, for any $s \geq 0$, it holds that $D_E(s\cdot\hat{\mathbf{\Lambda}}) = \overline{D}_E(s\cdot\hat{\Lambda})$, where D_E and $\overline{D}_E$ are defined in Eq.* (C.4).

Proof. It suffices to show that if $\sum_{y\in\mathcal{Y}_\ell} y^\top \hat{\Lambda}_y y = f(\hat{\Lambda})$, then $\sum_{y\in\mathcal{Y}_\ell} y^\top (s\cdot\hat{\Lambda}_y) y = f(s\cdot\hat{\Lambda})$ for any $s > 0$. By definition, we have

$$\begin{aligned}
f(s\cdot\hat{\Lambda}) = \max_{\Lambda_y} \quad & \textstyle\sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda_y y \\
\text{subject to} \quad & \textstyle\sum_{y\in\mathcal{Y}_\ell}\Lambda_y = s\cdot\hat{\Lambda} \\
& \Lambda_y \succeq \mathbf{0}, \quad \forall y \in \mathcal{Y}_\ell.
\end{aligned}$$

For the above optimization problem, we can do a change of variable by setting $\Lambda'_y = \frac{1}{s}\cdot\Lambda_y \implies \Lambda_y = s\cdot\Lambda'_y$. Then, we have

$$\begin{aligned}
f(s\cdot\hat{\Lambda}) = \max_{\Lambda_y} \quad & \textstyle\sum_{y\in\mathcal{Y}_\ell} y^\top (s\cdot\Lambda'_y) y \\
\text{subject to} \quad & \textstyle\sum_{y\in\mathcal{Y}_\ell} s\cdot\Lambda'_y = s\cdot\hat{\Lambda} \\
& s\cdot\Lambda'_y \succeq \mathbf{0}, \quad \forall y \in \mathcal{Y}_\ell.
\end{aligned}$$

$$\implies f(s\cdot\hat{\Lambda}) = \max_{\Lambda_y} \quad s\sum_{y\in\mathcal{Y}_\ell} y^\top \Lambda'_y y$$
$$\text{subject to} \quad \sum_{y\in\mathcal{Y}_\ell}\Lambda'_y = \hat{\Lambda}$$
$$\Lambda'_y \succeq \mathbf{0}, \quad \forall y\in\mathcal{Y}_\ell.$$

$$\implies f(s\cdot\hat{\Lambda}) = s\cdot f(\hat{\Lambda}) = s\cdot\sum_{y\in\mathcal{Y}_\ell} y^\top\Lambda_y y = \sum_{y\in\mathcal{Y}_\ell} y^\top(s\cdot\hat{\Lambda}_y)y.$$

Thus, the proof is complete. □

Lemma 45. *Let $f:\mathbb{R}^d\mapsto\mathbb{R}$ be a concave function with maximizer x^* over the convex set $\mathcal{C}$. Further, assume that f is G-strongly concave with respect to ℓ_2 norm in region $\mathcal{S}\cap\mathcal{C}$, where $\mathcal{S}=\{x:\|x-x^*\|_2\le A\}$. If $f(x^*)-f(x)\le\frac{AG}{2}$ and $c\in\mathcal{C}$, then $x\in\mathcal{S}$.*

Proof. By property of strong concavity, we know that, $f(x^*)-f(x)\ge\frac{G}{2}\|x-x^*\|_2$ for any $x\in\mathcal{S}\cap\mathcal{C}$. Now, suppose x' satisfies $f(x^*)-f(x')\le\frac{AG}{2}$, $x'\in\mathcal{C}$ and $x'\notin\mathcal{S}$. Then, we must have $\|x'-x^*\|_2>A$.

Let $\gamma\in(0,1)$ be some number such that $z=\gamma x'+(1-\gamma)x^*$ lies on the boundary of $\mathcal{S}$. By convexity, we also have $z\in\mathcal{C}$. Then, since f is concave, we have $f(z)\ge\gamma f(x')+(1-\gamma)f(x^*)>f(x')$, where the second inequality is strict because f is strongly concave in a region around x^*. Since $f(x^*)-f(x')\le\frac{AG}{2}$, f is G-strongly concave on $\mathcal{S}$ and z lies on the boundary of $\mathcal{S}$, we have

$$\frac{AG}{2}=\frac{G}{2}\|z-x^*\|_2\le f(x^*)-f(z)<f(x^*)-f(x')\le\frac{AG}{2}.$$

This is a contradiction and thus we must have $x'\in\mathcal{S}$. □

The following lemma quantitatively describes how close $\widetilde{\Lambda}$ and Λ^* needs to be to ensure an at most ϵ multiplicative constraint violation.

Lemma 46. *Assume $\|x\|_2\le M$ for any $x\in\mathcal{X}$. Let $\Sigma=\mathbb{E}_{X\sim\nu}\left[XX^\top\right]\succ\mathbf{0}$ and $\Lambda^*=\arg\max_{\Lambda\succeq\mathbf{0}}\overline{D}(\Lambda)$. Then, for any $\epsilon>0$, if we have*

$$\left\|\widetilde{\Lambda}-\Lambda^*\right\|_F\le\frac{8\mu^2\lambda_{\min}(\Sigma)}{3M^2\lambda_{\max}(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon},$$

then it holds that $y^\top\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)XX^\top\right]^{-1}y\le(1+\epsilon)c_\ell^2$ for any $y\in\mathcal{Y}_\ell$.

Proof. Fix some $\epsilon>0$. First, notice that if we regard P_Λ as a function of $q_\Lambda(x)=x^\top\Lambda x-1$, it then holds that

$$\|\nabla_\Lambda P_\Lambda(x)\|_2=\left\|\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}\nabla_\Lambda q_\Lambda(x)\right\|_2\le\left|\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}\right|\left\|xx^\top\right\|_2\le\left|\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}\right|M^2\le\frac{M^2}{8\mu},$$

where we obtain the last inequality by using Lemma 48. Therefore, for any $x\in\mathcal{X}$ and $\widetilde{\Lambda}\succeq\mathbf{0}$, we have $\left|P_{\widetilde{\Lambda}}(x)-P_{\Lambda^*}(x)\right|\le\frac{M^2}{8\mu}\cdot\left\|\widetilde{\Lambda}-\Lambda^*\right\|_F$ by mean value theorem and Cauchy-Schwartz. inequality.

Therefore, if we have $\left\|\widetilde{\Lambda}-\Lambda^*\right\|_F\le\delta$, then

$$\left|P_{\widetilde{\Lambda}}(x)-P_{\Lambda^*}(x)\right|\le\frac{M^2\delta}{8\mu}\implies P_{\widetilde{\Lambda}}(x)\ge P_{\Lambda^*}(x)-\frac{M^2\delta}{8\mu}$$

$$\implies\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)XX^\top\right]\succeq\mathbb{E}_{X\sim\nu}\left[P_{\Lambda^*}(X)XX^\top\right]-\frac{M^2\delta}{8\mu}\mathbb{E}_{X\sim\nu}\left[XX^\top\right].$$

By Lemma 47, we know that

$$y^\top \mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)XX^\top\right]^{-1} y \le c_\ell^2(1+\epsilon) \Longleftrightarrow \mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)XX^\top\right] \succeq \frac{yy^\top}{(1+\epsilon)c_\ell^2}. \tag{C.17}$$

Let $\Sigma^* = \mathbb{E}_{X\sim\nu}\left[P_{\Lambda^*}(X)XX^\top\right]$. Therefore, to guarantee the condition in Eq. (C.17), it is sufficient to guarantee that $\Sigma^* - \frac{M^2\delta}{8\mu}\Sigma \succeq \frac{yy^\top}{(1+\epsilon)c_\ell^2}$, which is equivalent to

$$w^\top\Sigma^* w - \frac{M^2\delta}{8\mu} w^\top\Sigma w \ge \frac{(w^\top y)^2}{c_\ell^2(1+\epsilon)}, \quad \forall \text{unit vector } w\in\mathbb{R}^d$$

$$\Longleftrightarrow \frac{1}{w^\top\Sigma w}\cdot w^\top\left(\Sigma^* - \frac{yy^\top}{(1+\epsilon)c_\ell^2}\right)w \ge \frac{M^2\delta}{8\mu}, \quad \forall \text{unit vector } w\in\mathbb{R}^d.$$

Therefore, it is sufficient to choose δ such that

$$\frac{M^2\delta}{8\mu} \le \frac{1}{\lambda_{\max}(\Sigma)}\cdot\lambda_{\min}\left(\Sigma^* - \frac{yy^\top}{c_\ell^2(1+\epsilon)}\right) \le \min_{w:\|w\|_2=1}\frac{1}{w^\top\Sigma w}\cdot w^\top\left(\Sigma^* - \frac{yy^\top}{(1+\epsilon)c_\ell^2}\right)w.$$

Since P_{Λ^*} satisfies the constraint defined in problem (C.5), we have $\Sigma^* \succeq \frac{yy^\top}{c_\ell^2}$. Meanwhile, by Lemma 48, we know that $P_{\Lambda^*}(x) \ge \frac{\mu}{3}$ for any $x\in\mathcal{X}$, which means that $\Sigma^* \succeq \frac{\mu}{3}\cdot\Sigma$. That is, for any unit vector $w\in\mathbb{R}^d$, we have

$$w^\top\Sigma^* w \ge \frac{\left(w^\top y\right)^2}{c_\ell^2} \quad \text{and} \quad w^\top\Sigma^* w \ge \frac{\mu}{3}\lambda_{\min}\left(\Sigma\right),$$

which together implies $w^\top\Sigma^* w \ge \max\left\{\frac{\mu}{3}\cdot\lambda_{\min}(\Sigma), \frac{\left(w^\top y\right)^2}{c_\ell^2}\right\}$. Therefore, it holds that

$$\begin{aligned}
w^\top\Sigma w - \frac{\left(w^\top y\right)^2}{(1+\epsilon)c_\ell^2} &\ge \max\left\{\frac{\mu}{3}\cdot\lambda_{\min}(\Sigma), \frac{\left(w^\top y\right)^2}{c_\ell^2}\right\} - \frac{\left(w^\top y\right)^2}{(1+\epsilon)c_\ell^2} \\
&= \max\left\{\frac{\mu}{3}\cdot\lambda_{\min}(\Sigma) - \frac{\left(w^\top y\right)^2}{(1+\epsilon)c_\ell^2}, \frac{\epsilon\left(w^\top y\right)^2}{(1+\epsilon)c_\ell^2}\right\} \\
&\ge \frac{\epsilon\mu}{3(1+\epsilon)}\cdot\lambda_{\min}\left(\Sigma\right) \\
\Longrightarrow\ \lambda_{\min}\left(\Sigma^* - \frac{yy^\top}{c_\ell^2(1+\epsilon)}\right) &\ge \frac{\epsilon\mu}{3(1+\epsilon)}\cdot\lambda_{\min}\left(\Sigma\right).
\end{aligned}$$

Therefore, to guarantee the condition in Eq. (C.17), it is sufficient to have

$$\frac{M^2\delta}{8\mu} = \frac{\epsilon\mu\lambda_{\min}(\Sigma)}{3(1+\epsilon)\lambda_{\max}(\Sigma)} \implies \mu = \frac{8\mu^2\lambda_{\min}(\Sigma)}{3M^2\lambda_{\max}(\Sigma)}\cdot\frac{\epsilon}{1+\epsilon},$$

Thus, the proof is complete. □

The following lemma is a result of standard Schur complement technique.

Lemma 47. *If $\mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right]$ is invertible and $c_\ell > 0$, then*

$$y^\top \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right]^{-1} y \le c_\ell^2 \iff \mathbb{E}_{X\sim\nu}\left[P(X)XX^\top\right] \succeq \frac{yy^\top}{c_\ell^2}.$$

Proof. For simplicity, let $A = E_{X\sim\nu}\left[P(X)XX^\top\right] \succ \mathbf{0}$. Then, we consider the block matrix $\begin{bmatrix} A & y \\ y^\top & c_\ell^2 \end{bmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$. Let $\begin{bmatrix} u & a \end{bmatrix}^\top \in \mathbb{R}^{d+1}$ with $u \in \mathbb{R}^d$ be some vector.

Now, for one direction, suppose $y^\top A^{-1} y \le c_\ell^2$ holds. Consider

$$\begin{bmatrix} u & a \end{bmatrix} \begin{bmatrix} A & y \\ y^\top & c_\ell^2 \end{bmatrix} \begin{bmatrix} u \\ a \end{bmatrix} = u^\top A u + 2a u^\top y + 2c_\ell^2 a^2 := r(u,a).$$

If we minimize $r(u,a)$ over u, which means to treat a as fixed, we can get (by taking gradient and setting it to zero)

$$u^* = -aA^{-1}y \implies r(u^*,a) = a^2(c_\ell^2 - y^\top A^{-1} y).$$

Since $y^\top A^{-1} y \le c_\ell^2$, we know that $r(u^*,a) \ge 0$, which means $r(u,a) \ge 0$ for any $\begin{bmatrix} u & a \end{bmatrix}^\top \in \mathbb{R}^{d+1}$.

Then, if we minimize $r(u,a)$ over a, we can get

$$a^* = -\frac{u^\top y}{c_\ell^2} \implies r(u,a^*) = u^\top A u - \frac{\left(u^\top y\right)^2}{c_\ell^2}.$$

Since $r(u,a) \ge 0$ for any $\begin{bmatrix} u & a \end{bmatrix}^\top \in \mathbb{R}^{d+1}$, we know that $u^\top A u - \frac{\left(u^\top y\right)^2}{c_\ell^2} \ge 0$ for any $u \in \mathbb{R}^d$. That is, we have $A \succeq \frac{yy^\top}{c_\ell^2}$.

The other direction simply takes the above calculation in a reversed way and thus the proof is complete. □

Properties of P_Λ

A visualization of P_Λ is given in Figure C.1.

Lemma 48. *The function $P_\Lambda(x)$ defined in* (4.7)*, if regarding as a function of $q_\Lambda(x) = x^\top \Lambda x - 1 \ge -1$, satisfies*

- $\lim_{q_\Lambda \to 0} P_\Lambda = \frac{1}{2}$ *for any* $\mu \in (0,1)$
- *When* $q_\Lambda = -1$, $P_\Lambda = \frac{1}{2} + \mu - \frac{\sqrt{1+4\mu^2}}{2} \ge \frac{\mu}{3}$ *and* $P_\Lambda - \mu(\log(1-P_\Lambda) + \log(P_\Lambda)) \le 2\sqrt{\mu}$ *for any* $\mu \in (0,1)$.
- $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda} = \frac{\mu\sqrt{q_\Lambda^2+4\mu^2}-2\mu^2}{q_\Lambda^2\sqrt{q_\Lambda^2+4\mu^2}}$ *decreases as q_Λ^2 increases. Further, $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda} \in [0, \frac{1}{8\mu}]$. Thus, P_Λ increases monotonically as q_Λ increases and $P_\Lambda(x) \ge \frac{\mu}{3}$ for any $x \in \mathcal{X}$ and $\Lambda \succeq \mathbf{0}$.*
- $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}|_{q_\Lambda=\pm 1} \ge \frac{\mu}{10}$ *and* $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda} \ge \frac{\mu}{2q_\Lambda^2}$ *when* $q_\Lambda^2 \ge 12\mu^2$.

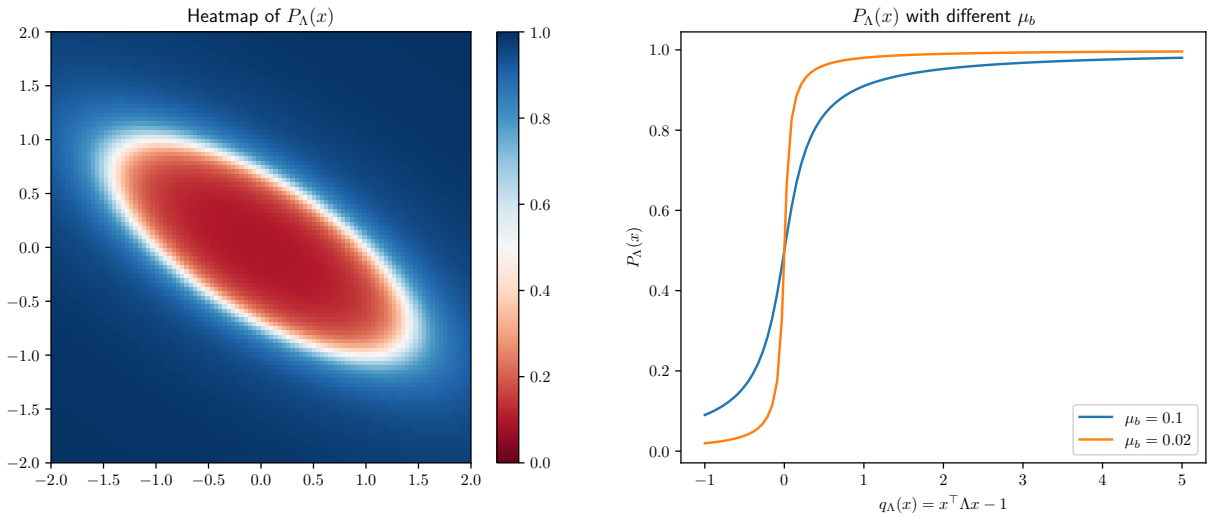


Figure C.1: (left) A heatmap of some P_Λ when problem dimension is $d = 2$, which shows that P_Λ is approximately an 0-1 threshold rule characterized by an ellipsoid. (right) A plot of P_Λ as a function of $q_\Lambda(x) = x^\top \Lambda x - 1$, which shows that the change of P_Λ near the boundary of ellipsoid is sharper when the barrier weight μ is smaller.

Proof. For simplicity, we will drop the subscript Λ and just treat P as a function of q. That is, we have

$$P(q) = \frac{1}{2} - \frac{\mu}{q} + \frac{\sqrt{(2\mu - q)^2 + 4\mu q}}{2q}.$$

We prove each bullet point separately.

- Since $P(q)$ also satisfies Eq. (4.6), which in simpler form is $\frac{\mu}{1-P(q)} - \frac{\mu}{P(q)} = q$, we can easily see that $P(q) = \frac{1}{2}$ satisfies this equation when $q = 0$.

- By direction computation, we can get $P(-1) = \frac{1}{2} + \mu - \frac{\sqrt{1+4\mu^2}}{2}$. To show this is greater than $\frac{\mu}{3}$ for any $\mu \in [0, 1]$, consider $\ell(\mu) = P(-1) - \frac{\mu}{3}$. It is easy to check that $\ell(0) = 0$ and $\ell(1) > 0$. Then, since $\ell'(\mu) = \frac{2}{3} - \frac{2\mu}{\sqrt{1+4\mu^4}}$ is initially greater than 0 and then smaller than 0, we know $\ell(\mu)$ first increases and then decreases on $[0, 1]$. Thus, $\ell(\mu) \geq 0$ on $[0, 1]$ and thus $P(-1) \geq \frac{\mu}{3}$ for any $\mu \in [0, 1]$.

  For the second part, define $\tilde{\ell}(\mu) = 2\sqrt{\mu} - P(-1) + \mu\left(\log(1 - P(-1)) + \log(P(-1))\right)$. Then, by utilizing the fact that P satisfies Eq. (4.6), we can compute its derivative and get $\frac{\mathrm{d}\tilde{\ell}}{\mathrm{d}\mu} = \frac{1}{\sqrt{\mu}} + \log(1 - P(-1)) + \log(P(-1))$. We can check that on the domain $(0, 1)$, we have $\frac{\mathrm{d}^2\tilde{\ell}}{\mathrm{d}\mu^2} = -\frac{1}{2\mu^{3/2}} + \frac{1}{\mu} - \frac{2}{\sqrt{1+4\mu^2}} \cdot \frac{2\sqrt{\mu(1+4\mu^2)} - 4\mu^{3/2} - \sqrt{1+4\mu^2}}{2\mu^{3/2}\sqrt{1+4\mu^2}} \leq 0$ on $(0, 1)$, which means that $\frac{\mathrm{d}\tilde{\ell}}{\mathrm{d}\mu}$ is monotonically decreasing. To see why the second derivative is smaller than 0, we can compute

  $$\left(4\mu^{3/2} + \sqrt{1 + 4\mu^2}\right) - 4\mu\left(1 + 4\mu^2\right) = (1 - 2\mu)^2 + 8\mu^{3/2}\sqrt{1 + 4\mu^2} \geq 0.$$

  Thus, $\frac{\mathrm{d}\tilde{\ell}}{\mathrm{d}\mu}$ is initially greater than 0 and then smaller than 0 on $(0, 1)$. It is easy to verify that $\lim_{\mu \to 0} \tilde{\ell} = 0$ and $\tilde{\ell}(1) > 0$. Therefore, we have $\tilde{\ell}(\mu) \geq 0$ for any $\mu \in (0, 1)$.

- We can get $\frac{\mathrm{d}P}{\mathrm{d}q} = \frac{\mu\sqrt{q^2+4\mu^2}-2\mu^2}{q^2\sqrt{q^2+4\mu^2}}$ by direct computation. To show it is decreasing as q^2 increasing, we consider $\tilde{f}(z) = \frac{\mu\sqrt{z+4\mu^2}-2\mu^2}{z\sqrt{z+4\mu^2}}$ and it is sufficient to show that $\frac{\mathrm{d}\tilde{f}}{\mathrm{d}z} < 0$ for any $z > 0$. Again by direct computation, we have

$$\frac{\mathrm{d}\tilde{f}}{\mathrm{d}z} = \frac{\mu\left(8\mu^3 + 3\mu z - \left(z+4\mu^2\right)^{3/2}\right)}{z^2\left(z+4\mu^2\right)^{3/2}}.$$

  By direct computation, We can show that $\left(z+4\mu^2\right)^3 - \left(8\mu^3+3\mu z\right)^2 = z^3 + 3z^2\mu^2 > 0$ for any $z > 0$ and $\mu \in [0,1]$. Thus, $\frac{\mathrm{d}\tilde{f}}{\mathrm{d}z} < 0$ and thus $\frac{\mathrm{d}P}{\mathrm{d}q}$ is decreasing as q^2 increases.

  It is obvious that $\frac{\mathrm{d}P}{\mathrm{d}q} \geq 0$ for any $q^2 \geq 0$ and $\mu \in [0,1]$ since we always have $\mu\sqrt{q^2+4\mu^2} \geq 2\mu^2$. Thus, the maximum value could potentially happen is when $q^2 \to 0$, which can be evaluated by using L'Hospital's rule. A direct computation gives us $\lim_{q^2\to 0}\frac{\mathrm{d}P}{\mathrm{d}q} = \frac{1}{8\mu}$. Thus, we can conclude that $\frac{\mathrm{d}P}{\mathrm{d}q} \in \left[0, \frac{1}{8\mu}\right]$. Therefore, P increases monotonically as q increases, which implies that $P_\Lambda(x) \geq \frac{\mu}{3}$ for any $x \in \mathcal{X}$ and Λ.

- By direct computation, we have $\frac{\mathrm{d}P_\Lambda}{\mathrm{d}q_\Lambda}|_{q_\Lambda=\pm 1} = \mu\left(1 - \frac{2\mu}{\sqrt{1+4\mu^2}}\right) \geq \mu\left(1-\frac{2}{\sqrt{5}}\right) \geq \frac{\mu}{10}$ for any $\mu \in [0,1]$. The reason is that we can easily see $\frac{2\mu}{\sqrt{1+4\mu^2}}$ is increasing in μ.

  Finally, notice that when $2\mu \leq \frac{1}{2}\sqrt{q^2+4\mu^2}$, which is equivalent to $q^2 \geq 12\mu^2$, we have

$$\frac{\mathrm{d}P}{\mathrm{d}q} = \frac{\mu\sqrt{q^2+4\mu^2}-2\mu^2}{q^2\sqrt{q^2+4\mu^2}} \geq \frac{\mu\sqrt{q^2+4\mu^2}-\frac{\mu}{2}\sqrt{q^2+4\mu^2}}{q^2\sqrt{q^2+4\mu^2}} = \frac{\mu}{2q^2}.$$

Thus, the proof is complete. □

C.3.3 An Alternative Approach to OPTIMIZEDESIGN

Based on the analysis in Section C.3.1, we know that maximizing $\overline{D}(\cdot)$ is equivalent to maximizing $D(\cdot)$. Therefore, as an alternative to Algorithm 4.2, which maximizes $D(\cdot)$ through stochastic gradient ascent, it is natural to have an algorithm that directly maximizes $\overline{D}(\cdot)$. Here, we will consider subgradient ascent.

Recall that $\overline{D} : \mathbb{S}_+^d \mapsto \mathbb{R}$ is defined as

$$\overline{D}(\Lambda) = \mathbb{E}_{X\sim\nu}\left[P_\Lambda(X) - \mu\left(\log(1-P_\Lambda(X)) + \log(P_\Lambda(X))\right) - P_\Lambda(X)X^\top\Lambda X\right] + \frac{1}{c_\ell^2}\cdot f(\Lambda),$$

where $f(\Lambda)$ is defined in problem (C.3). The subgradient of $\overline{D}(\Lambda)$ is

$$\partial\overline{D}(\Lambda) = \mathbb{E}_{X\sim\nu}\left[\left(1+\frac{\mu}{1-P_\Lambda(x)} - \frac{\mu}{P_\Lambda(X)} - X^\top\Lambda X\right)\nabla P_\Lambda(X) - P_\Lambda(X)XX^\top\right] + \frac{\partial f(\Lambda)}{c_\ell^2}$$

(The first term is differentiable)

$$= \frac{\partial f(\Lambda)}{c_\ell^2} - \mathbb{E}_{X\sim\nu}\left[P_\Lambda(X)XX^\top\right].$$

(Since $P_\Lambda(X)$ solves Eq. (4.6))

Therefore, to run subgradient ascent, we only need to find an element in $\partial f(\Lambda)$, which can be obtained by solving the following optimization problem as claimed by Lemma 49.

$$\begin{array}{rl} \min_{\Gamma} & \langle \Gamma, \Lambda \rangle \\ \text{subject to} & \Gamma \succeq yy^\top, \quad \forall y \in \mathcal{Y}_\ell, \\ & \Gamma \preceq 2\sum_{y \in \mathcal{Y}_\ell} yy^\top. \end{array} \tag{C.18}$$

As a result, we have Algorithm C.1 as an alternative to solve OPTIMIZEDESIGN. Compared to Algorithm 4.2, which needs to maintain $|\mathcal{Y}_\ell|\, d^2$ number of objective variables, Algorithm C.1 only has d^2 variables. However, each iteration of Algorithm C.1 is computationally more intensive since finding a subgradient needs to solve the problem (C.18).

Algorithm C.1. Projected Stochastic Subgradient Ascent to Solve OPTIMIZEDESIGN

Input: Number of iterations K; number of samples u; barrier weight $\mu_b \in (0,1)$
Initialize $\hat{\Lambda}^{(0)} = \mathbf{0}$
for $k = 0, 1, 2, \ldots, K-1$ **do**
  Sample $x_k \sim \nu$
  Solve problem (C.18) with current $\hat{\Lambda}^{(k)}$ to get $\Gamma^{(k)}$
  Set $g_k = \frac{\Gamma^{(k)}}{c_\ell^2} - P_{\hat{\Lambda}^{(k)}}(x_k) x_k x_k^\top$
  Set $\hat{\Lambda}^{(k+1)} \leftarrow \hat{\Lambda}^{(k)} + \eta_k g_k$, where $\eta_k = \frac{1}{\sqrt{2\sum_{s=1}^k \|g_s\|_2^2}}$
  Update $\hat{\Lambda}^{(k+1)} \leftarrow \Pi_{\mathbb{S}_+^d}(\hat{\Lambda}^{(k+1)})$, a projection to the set of $d \times d$ PSD matrices
Let $\hat{\Lambda} = \frac{1}{K}\sum_{k=1}^K \hat{\Lambda}^{(k)}$
Find $s^* \leftarrow \arg\max_{s \in [0,1]} \overline{D}_E(s \cdot \hat{\Lambda})$, where $\overline{D}_E$ is the empirical version of $\overline{D}$, evaluated using u i.i.d. samples
return $\widetilde{\Lambda} = s^* \cdot \hat{\Lambda}$

A result similar to Theorem 23 can also be obtained for Algorithm C.1, which is given in Theorem 24. The bounds are almost identical except that the old lower bound for K depends on $|\mathcal{Y}_\ell|^3$ while the new one depends on $|\mathcal{Y}_\ell|$. Steps identical to the proof of Theorem 23 will be skipped in the proof of Theorem 24.

Theorem 24. *Let $\Lambda^* \in \arg\max_{\Lambda \succeq \mathbf{0}} \overline{D}(\Lambda)$ and take other settings the same as that in Theorem 23. Then, Λ^* is unique. Further, for any $\epsilon > 0$ and $\delta > 0$, suppose it holds that*

$$\begin{aligned} \mu &\le \min\left\{ \sqrt{\frac{3\kappa(\Sigma)\,\|\Lambda^*\|_F\, M^2}{8} \cdot \frac{1+\epsilon}{\epsilon}}, \frac{4}{9}\|\Lambda^*\|_F^2\, M^4, \frac{1}{2\sqrt{3}} \right\} \\ K &\ge \frac{288\kappa(\Sigma)^2\,\|\Lambda^*\|_F^4\, M^4 (M^4 + 4\,|\mathcal{Y}_\ell|\, C_\ell^2) \cdot \left(2\,\|\Lambda^*\|_F\, M^2 + 1\right)^4 \log(6/\delta)}{\omega^2 \mu^6} \cdot \left(\frac{1+\epsilon}{\epsilon}\right)^2 \\ u &\ge \frac{576\kappa(\Sigma)^2\,\|\Lambda^*\|_F^2\, M^8 \cdot \left(2\,\|\Lambda^*\|_F\, M^2 + 1\right)^4 \log(6/\delta)}{\omega^2 \mu^6} \cdot \left(\frac{1+\epsilon}{\epsilon}\right)^2. \end{aligned}$$

Then, with probability at least $1-\delta$, Algorithm 4.2 will output $\widetilde{\Lambda}$ that satisfies

- $y^\top \mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X) X X^\top\right]^{-1} y \le (1+\epsilon) c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell.$
- $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right] \le \mathbb{E}_{X\sim\nu}\left[\widetilde{P}(X)\right] + 4\sqrt{\mu}$, *where $\widetilde{P}$ is the optimal solution to problem* (C.13).

Proof. **First Bullet Point.** Similar to the proof of Theorem 23, let $\hat{\Lambda}$ be the parameter obtained by Algorithm C.1 just before the re-scaling step (line 10). Then, by Theorem 3.13 of (Orabona, 2019), with probability at least $1 - \frac{\delta}{3}$, it holds that

$$\overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) \leq \frac{\mathrm{Reg}(K) + 2\sqrt{2K\log(6/\delta)}}{K},$$

where $\mathrm{Reg}(K)$ is the regret of running projected stochastic subgradient ascent for K steps with η_k specified in Algorithm C.1. Meanwhile, by Theorem 4.14 of (Orabona, 2019) also, we have $\mathrm{Reg}(K) = \sqrt{2}B^2\sqrt{\sum_{k=1}^K \|g_k\|_2^2}$, where $B = \|\Lambda^*\|_F$. Since $g_k = \frac{\Gamma^{(k)}}{c_\ell^2} - P_{\hat{\Lambda}^{(k)}}(x_k)x_k x_k^\top$ and $\left\|\Gamma^{(k)}\right\|_F \leq 2\left\|\sum_{y\in\mathcal{Y}_\ell} yy^\top\right\|_F$, we can easily get $\|g_k\|_2^2 \leq 2M^4 + \frac{8}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell}\|y\|_2^4 = 2M^4 + 8\left|\mathcal{Y}_\ell\right| C_\ell^2$. Thus, we have

$$\mathrm{Reg}(K) \leq 2\left\|\Lambda^*\right\|_F^2\sqrt{M^4 + 4\left|\mathcal{Y}_\ell\right| C_\ell^2}\cdot\sqrt{K} := C_{\mathrm{Reg}}\sqrt{K} \tag{C.19}$$

$$\implies \overline{D}(\Lambda^*) - \overline{D}(\hat{\Lambda}) \leq \frac{C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}}{\sqrt{K}}, \tag{C.20}$$

We now consider the effect of using u i.i.d. samples in the re-scaling step. Since re-scaling always increases the function value, we must have $\overline{D}_E(\hat{\Lambda}) \leq \overline{D}_E(\widetilde{\Lambda})$.

Then, after **exactly the same** steps of analysis, we can get the following same lower bound for K,

$$K \geq \left(\frac{3\kappa(\Sigma)M^2\left(C_{\mathrm{Reg}} + 2\sqrt{2\log(6/\delta)}\right)}{2G\mu^2}\cdot\frac{1+\epsilon}{\epsilon}\right)^2, \tag{C.21}$$

with a different value of C_{Reg}.

Second Bullet Point. We then prove the upper bound for primal objective value $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right]$, which explains the reason why an extra re-scaling step is needed. Let $\hat{\mathbf{\Lambda}} = (\hat{\Lambda}_y)_{y\in\mathcal{Y}_\ell}$ be a set of PSD matrices that solves problem (C.3) with parameter $\hat{\Lambda}$ and $\widetilde{\mathbf{\Lambda}} = s^*\cdot\hat{\mathbf{\Lambda}}$, where $s^* = \arg\max_{s\in[0,1]}\overline{D}_E(s\cdot\hat{\Lambda})$. Since the constraint in problem (C.3) requires $\sum_{y\in\mathcal{Y}_\ell}\hat{\Lambda}_y = \hat{\Lambda}$, we have $\sum_{y\in\mathcal{Y}_\ell}\widetilde{\Lambda}_y = \widetilde{\Lambda}$, which is the output of Algorithm C.1.

Define $g(s) = D_E(s\cdot\widetilde{\mathbf{\Lambda}})$. By construction, we know that $g(s)$ is maximized at $s = 1$ because $\overline{D}_E(s\cdot\hat{\Lambda}) = D_E(s\cdot\hat{\mathbf{\Lambda}})$ for any $s \geq 0$ as shown in Lemma 44, which means that $s^* = \arg\max_{s\in[0,1]} D_E(s\cdot\hat{\mathbf{\Lambda}})$. Therefore, we have $g'(1) \geq 0$, which in turn gives us

$$g'(1) = \frac{1}{c_\ell^2}\sum_{y\in\mathcal{Y}_\ell} y^\top\widetilde{\Lambda}_y y - \frac{1}{u}\sum_{i=1}^u P_{\widetilde{\Lambda}}(x_i)x_i^\top\widetilde{\Lambda}x_i \geq 0.$$

Then, after **exactly the same** steps of analysis, we can get $\mathbb{E}_{X\sim\nu}\left[P_{\widetilde{\Lambda}}(X)\right] \leq \mathbb{E}_{X\sim\nu}\left[\widetilde{P}(X)\right] + 4\sqrt{\mu}$, where $\widetilde{P}$ is the optimal solution of the problem (C.13). □

Technical Lemmas

Lemma 49. *The optimal value of the optimization problem* (C.18) *with parameter* $\Lambda \succeq \mathbf{0}$ *is equal to* $f(\Lambda)$. *Further, let* $\Gamma^*(\Lambda)$ *be an optimal solution to* (C.18). *Then, it holds that* $\Gamma^*(\Lambda) \in \partial f(\Lambda)$ *and* $\|\Gamma^*(\Lambda)\| \leq 2\left\|\sum_{y\in\mathcal{Y}_\ell} yy^\top\right\|_F$.

Proof. Alternatively, we first consider the following optimization problem.

$$\begin{array}{ll} \max_{\Lambda_y, \Sigma} & \sum_{y \in \mathcal{Y}_\ell} y^\top (\Lambda_y - 2\Sigma) y \\ \text{subject to} & \Lambda = \sum_{y \in \mathcal{Y}_\ell} \Lambda_y - \Sigma, \\ & \Sigma \succeq \mathbf{0}, \Lambda_y \succeq \mathbf{0}, \quad \forall y \in \mathcal{Y}_\ell. \end{array} \tag{C.22}$$

Since $y^\top \Sigma y \geq 0$ for any $y \in \mathcal{Y}_\ell$ and $\Sigma \succeq \mathbf{0}$, it is clear that problem (C.22) has the same optimal value as problem (C.3). Then, let $\Gamma \in \mathbb{R}^{d \times d}$ be the dual variable for the equality constraint in problem (C.22). We can have its dual problem to be

$$\begin{aligned} & \min_{\Gamma} \max_{\substack{\Lambda_y \succeq \mathbf{0}, \forall y \in \mathcal{Y}_\ell, \\ \Sigma \succeq \mathbf{0}}} \sum_{y \in \mathcal{Y}_\ell} \left\langle yy^\top, \Lambda_y - 2\Sigma \right\rangle + \left\langle \Gamma, \Lambda + \Sigma - \sum_{y \in \mathcal{Y}_\ell} \Lambda_y \right\rangle \\ \Longrightarrow & \min_{\Gamma} \max_{\substack{\Lambda_y \succeq \mathbf{0}, \forall y \in \mathcal{Y}_\ell, \\ \Sigma \succeq \mathbf{0}}} \langle \Gamma, \Lambda \rangle + \left\langle \Sigma, \Gamma - 2 \sum_{y \in \mathcal{Y}_\ell} yy^\top \right\rangle + \sum_{y \in \mathcal{Y}_\ell} \left\langle \Lambda_y, yy^\top - \Gamma \right\rangle. \end{aligned}$$

In order for the above optimization problem to have finite value, we must have $\Gamma \preceq 2 \sum_{y \in \mathcal{Y}_\ell} yy^\top$ and $\Gamma \succeq yy^\top$ for any $y \in \mathcal{Y}_\ell$. Therefore, we obtain the following dual problem.

$$\begin{array}{ll} \min_{\Gamma} & \langle \Gamma, \Lambda \rangle \\ \text{subject to} & \Gamma \succeq yy^\top, \quad \forall y \in \mathcal{Y}_\ell, \ . \\ & \Gamma \preceq 2 \sum_{y \in \mathcal{Y}_\ell} yy^\top. \end{array}$$

This is exactly the problem (C.18). Then, we can notice the Slater's condition is clearly satisfied by problem (C.18), which means the strong duality holds. Therefore, problem (C.18) has the same optimal value as (C.22), which is the same as (C.3).

Since $f(\Lambda)$ is concave in Λ as shown in Lemma 40, to show that $\Gamma^*(\Lambda) \in \partial f(\Lambda)$, consider arbitrary $\Lambda, \Lambda' \succeq \mathbf{0}$. Then, we have

$$f(\Lambda) + \left\langle \Gamma^*(\Lambda), \Lambda' - \Lambda \right\rangle = \langle \Gamma^*(\Lambda), \Lambda \rangle + \left\langle \Gamma^*(\Lambda), \Lambda' - \Lambda \right\rangle = \left\langle \Gamma^*(\Lambda), \Lambda' \right\rangle \geq f(\Lambda').$$

The first equality holds because the optimal value of problem (C.18) is $f(\Lambda)$ as just shown above. The last inequality holds because $\Gamma^*(\Lambda)$ is a feasible solution to the problem (C.18) with parameter Λ'. Therefore, we have $\Gamma^*(\Lambda) \in \partial f(\Lambda)$.

Finally, since the constraint of problem (C.18) requires $\Gamma^*(\Lambda) \preceq 2 \sum_{y \in \mathcal{Y}_\ell} yy^\top$, we can obtain $\|\Gamma^*(\Lambda)\|_F \leq 2 \left\| \sum_{y \in \mathcal{Y}_\ell} yy^\top \right\|_F$ as a direct consequence of Lemma 50. □

Lemma 50. *For $A, B \in \mathbb{S}^{d \times d}$, if $A \succeq B \succeq \mathbf{0}$, then $\|A\|_F \geq \|B\|_F$.*

Proof. Let $\lambda_1, \ldots, \lambda_d$ and $\gamma_1, \ldots, \gamma_d$ be eigenvalues of A and B, respectively. Let $v_1, \ldots, v_d$ be a set of orthogonal unit eigenvectors of matrix A. Then, we have

$$\|A\|_F = \sqrt{\operatorname{tr}(AA)} = \sqrt{\operatorname{tr}\left(\left(\sum_{i=1}^d \lambda_i v_i v_i^\top \right) \left(\sum_{i=1}^d \lambda_i v_i v_i^\top \right) \right)} = \sqrt{\sum_{i=1}^d \lambda_i^2}.$$

Similarly, we have $\|B\|_F = \sqrt{\sum_{i=1}^d \gamma_i^2}$. By Corollary 7.7.4 in (Horn and Johnson, 2012), since $A \succeq B \succeq \mathbf{0}$, we know that $\lambda_i \geq \gamma_i \geq 0$ for each i. Therefore, we have $\|A\|_F \geq \|B\|_F$. □

C.4 Selective Sampling Algorithm for Unknown Distribution ν

C.4.1 Statement and proof of Theorem 25

Consider now the case where we do not know ν exactly, and are returned $(\widehat{P}_\ell, \widehat{\Sigma}_{\widehat{P}_\ell})$ that only approximate their ideals. Algorithm 4.1 can still be employed to solve this case where ν is unknown, but at the cost of sampling some historical data. Note that compared to the case where ν is know, it assumes the knowledge of an upper bound on $\sup_{x\in\text{support}(\nu)} \|x\|$. It also relies on a multiplicative factor change in the constraint of the optimization problem, in order to account for the possible constraint violation of the output of the subroutine. The last difference is the use of an approximation of the covariance matrix to compute the estimator. The covariance matrix is empirically approximated by injecting additional unlabeled samples (historical data). With that, although we do not know ν but we can approximate the relevant quantities, such as the covariance matrix $\mathbb{E}_{X\sim\nu}[XX^\top]$.

Let us detail the properties of the implementation of $\widehat{P}_\ell, \widehat{\Sigma}_{\widehat{P}_\ell} \leftarrow$OPTIMIZEDESIGN$(\mathcal{Z}_\ell, 2^{-\ell}, \tau)$ we use at each round ℓ.

First, $\widehat{P}_\ell$ has the properties described in Theorem 13 (by using Algorithm 4.2). More explicitly, let $\epsilon_\ell := 2^{-\ell}$, $B < \infty$ such that $\max_{x\in\mathcal{X}} |\langle x, \theta_*\rangle| \le B$, and $\sigma < \infty$ such that $\mathbb{E}[(y_s - \langle\theta_*, x_s\rangle)^2 | x_s] \le \sigma^2$. If

$$\beta_{\delta,\ell} := 4(1+\varepsilon)^2 \left(4\sqrt{B^2+\sigma^2}+1\right)^2 \log(4\ell^2|\mathcal{Z}|^2/\delta)$$

then $\widehat{P}_\ell$ is such that

- $\max_{z,z'\in\mathcal{Z}_\ell} \frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau\widehat{P}_\ell(X)XX^\top]^{-1}}}{\epsilon_\ell^2}\beta_{\delta,\ell} \le 1+\varepsilon.$
- $\mathbb{E}_{X\sim\nu}\left[\widehat{P}_\ell(X)\right] \le \mathbb{E}_{X\sim\nu}\left[\widetilde{P}_\ell(X)\right] + 4\sqrt{\mu_b}$, where $\widetilde{P}_\ell$ is the optimal solution to problem (C.23).

$$\begin{array}{rl} \min_P & \mathbb{E}_{X\sim\nu}[P(X)] \\ \text{subject to} & \max_{z,z'\in\mathcal{Z}_\ell} \frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2}\beta_{\delta,\ell} \le 1, \\ & 0 \le P(x) \le 1-\mu_b, \quad \forall x\in\mathcal{X}. \end{array} \tag{C.23}$$

where $\mu_b \ge 0$. The quantity $\mathbb{E}_{X\sim\nu}\left[\widetilde{P}_\ell(X)\right]$ that uses $\mu_b > 0$ is easily related to the value when $\mu_b = 0$ through a simple scaling factor of $\frac{1}{1-\mu_b}$ (see proof below).

$\widehat{\Sigma}_{\widehat{P}_\ell}$ is the empirical covariance matrix of $\Sigma_{\widehat{P}_\ell} := \mathbb{E}_{X\sim\nu}[\widehat{P}_\ell(X)XX^\top]$ using historical data and is such that

$$(1-\gamma)\Sigma_{\widehat{P}_\ell} \preceq \widehat{\Sigma}_{\widehat{P}_\ell} \preceq (1+\gamma)\Sigma_{\widehat{P}_\ell}$$

where $\gamma \ge 0$.

Again, while we think of historical data as independent data collected offline before the start of the game, in practice this historical data could just come from previous rounds (which is not technically correct since its use may introduce some dependencies).

Theorem 25 (Upper bound)**.** *Fix any $\delta \in (0,1)$. Let $\Delta = \min_{z\in\mathcal{Z}\setminus z_*}\langle z_* - z, \theta_*\rangle$ and set*

$$\beta_\delta = 256(1+\varepsilon)^2\left(4\sqrt{B^2+\sigma^2}+1\right)^2 \log(4\log_2^2(\tfrac{4}{\Delta})|\mathcal{Z}|^2/\delta).$$

For any $\tau \geq \rho(\nu)\beta_\delta$ there exists a δ-PAC selective sampling algorithm that collects $\mathcal{T}$ historical data before the start of the game, observes $\mathcal{U}$ unlabeled examples, and requests just $\mathcal{L}$ labels that satisfies

- $\mathcal{U} \leq \log_2(\frac{4}{\Delta})\tau$,
- $\mathcal{L} \leq \frac{1}{1-\mu_b} \min_{\lambda \in \triangle_{\mathcal{X}}} \rho(\lambda)\beta_\delta + \frac{5\tau}{1-\mu_b}\sqrt{\mu_b}$ *subject to* $\tau \geq \|\lambda/\nu\|_\infty \rho(\lambda)\, \beta_\delta$, *and*
- $\mathcal{T} \leq \log_2(\frac{4}{\Delta})(K + u + \kappa_\delta)$

with probability at least $1-\delta$.

Here, the sample complexity for estimating the covariance matrix is bounded by $\kappa_\delta = \lceil 2K_{\psi_2}^2(\sqrt{d \ln 9/c_1} + \sqrt{\frac{\log(2/\delta)}{c_1}}) \max\{1, 20\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]}\}\rceil$ (where the sub-gaussian norm $K_{\psi_2} = \max_{s,P} \|\sqrt{P(\widetilde{x}_s)}\Sigma_P^{-1/2}\widetilde{x}_s\|_{\psi_2}$), and the contributions from the optimization problem to compute $\{\widehat{P}_\ell\}_\ell$ are

$$K = \widetilde{O}\left(\frac{|\mathcal{Z}|^6\, \kappa(\Sigma)^2\, \|\Lambda^*\|_2^8\, M^{16}}{\omega^2 \mu_b^6}\right) \cdot \left(\frac{1+\epsilon}{\epsilon}\right)^2, \quad u = \widetilde{O}\left(\frac{\kappa(\Sigma)^2\, \|\Lambda^*\|_2^6\, M^{16}}{\omega^2 \mu_b^6}\right) \cdot \left(\frac{1+\epsilon}{\epsilon}\right)^2,$$

Naturally, we have a trade-off on the subroutine tolerance μ_b. In order to get a better solution of the optimization over the selection rule P (and thus get a smaller $\sum_{t=(\ell-1)\tau+1}^{\ell\tau} P(x_t)$ term), the subroutine needs more unlabeled samples. However, it suffices to take $\mu_b = \frac{1}{\tau^2}$ to make $\mathcal{U}$, and $\mathcal{L}$ roughly match those of the case when ν was known.

The proof of this theorem is established through several results, which we provide in Section C.4.2.

C.4.2 Lemmas for the correctness

We first state here the correctness of Algorithm 4.1 in the case where ν is unknown.

Lemma 51. *With probability at least $1-\delta$ we have for all stages $\ell \in \mathbb{N}$, we have that $z_* \in \mathcal{Z}_\ell$ and $\max_{z\in\mathcal{Z}_\ell}\langle z_* - z, \theta_*\rangle \leq 4\epsilon_\ell$.*

The proof of the correctness lemma is established though several lemmas. First we provide Lemma 52 guaranteeing concentration of empirical covariance matrices, which is obtained by sampling κ additional measurements. Then we show in Proposition 5 that the RIPS estimator does not suffer from using that empirical covariance matrix.

Lemma 52. *For any $P : \mathcal{X} \to [0,1]$, let $\Sigma_P = \mathbb{E}_{X\sim\nu}[P(X)XX^\top]$, $\widehat{\Sigma}_P = \frac{1}{\kappa}\sum_{s=1}^{\kappa} P(\widetilde{x}_s)\widetilde{x}_s\widetilde{x}_s^\top$. Define $K_{\psi_2} = \max_s \|\sqrt{P(\widetilde{x}_i)}\Sigma_P^{-1/2}\widetilde{x}_s\|_{\psi_2}$. With probability at least $1 - 2\exp(-c_1 t^2/K_{\psi_2}^4)$ holds*

$$(1-c)x^\top\Sigma_P x \leq x^\top\widehat{\Sigma}_P x \leq (1+c)x^\top\Sigma_P x$$

where $c = \max\left\{\frac{C\sqrt{d}+t}{\sqrt{\kappa}}, \left(\frac{C\sqrt{d}+t}{\sqrt{\kappa}}\right)^2\right\}$, $C = K_{\psi_2}^2\sqrt{\ln 9/c_1}$ and c_1 is an absolute constant.

Consequently for $\kappa \geq c_\delta := K_{\psi_2}^2(\sqrt{d\ln 9/c_1} + \sqrt{\frac{\log(2/\delta)}{c_1}})$, holds with probability at least $1-\delta$

$$\left(1 - \frac{c_\delta}{\sqrt{\kappa}}\right) x^\top\Sigma_P x \leq x^\top\widehat{\Sigma}_P x \leq \left(1 + \frac{c_\delta}{\sqrt{\kappa}}\right) x^\top\Sigma_P x.$$

Proof. Let $A \in \mathbb{R}^{\kappa \times d}$ whose rows A_i are independent sub-gaussian isotropic random vectors in R^d and define $K_{\psi_2} = \max_i \|A_i\|_{\psi_2}$. We can apply Theorem 5.39 of (Vershynin, 2011) on A to have that with probability at least $1 - 2\exp(-c_1 t^2/K_{\psi_2}^4)$ holds

$$1 - \frac{C\sqrt{d}+t}{\sqrt{\kappa}} \leq \sigma_{\min}(A) \leq \sigma_{\max}(A) \leq 1 + \frac{C\sqrt{d}+t}{\sqrt{\kappa}},$$

where $C = K_{\psi_2}^2\sqrt{\ln 9/c_1}$ and c_1 is an absolute constant.

With Lemma 5.36 of (Vershynin, 2011), this implies that with probability at least $1-2\exp(-c_0 t^2)$ holds

$$\|A^\top A - I\| \leq \max\left\{\frac{C\sqrt{d}+t}{\sqrt{\kappa}}, \left(\frac{C\sqrt{d}+t}{\sqrt{\kappa}}\right)^2\right\} =: c \tag{C.24}$$

Recall $\Sigma_P = \mathbb{E}_{X\sim\nu}[P(X)XX^\top]$, so $Y = \sqrt{P(X)}\Sigma_P^{-1/2}X$ satisfies $\mathbb{E}[YY^\top] = \mathbb{E}[\Sigma_P^{-1/2}P(X)XX^\top\Sigma_P^{-1/2}] = \Sigma_P^{-1/2}\Sigma_P\Sigma_P^{-1/2} = I$. So we can apply (C.24) to get $\|\Sigma_P^{-1/2}\widehat{\Sigma}_P\Sigma_P^{-1/2} - I\| \leq c$. Thus for any $y \in \mathbb{R}^d$,

$$1 - c \leq \frac{y^\top}{\|y\|}\Sigma_P^{-1/2}\widehat{\Sigma}_P\Sigma_P^{-1/2}\frac{y}{\|y\|} \leq 1 + c$$

so setting $y = \Sigma_P^{1/2}x$

$$(1-c)x^\top\Sigma_P x \leq x^\top\widehat{\Sigma}_P x \leq (1+c)x^\top\Sigma_P x.$$

Also, the sub-gaussian bound becomes $K_{\psi_2} = \max_i \|\sqrt{P(\widetilde{x}_i)}\Sigma_P^{-1/2}\widetilde{x}_i\|_{\psi_2}$. □

Proposition 5 (RIPS guarantees on empirical covariance matrix)**.** *Let $x_1, \ldots, x_n$ and $\widetilde{x}_1, \ldots, \widetilde{x}_\kappa$ be drawn IID from a distribution ν. For $s = 1, \ldots, n$, assume that $|\langle\theta, x_s\rangle| \leq B$ and $\mathbb{E}[|\langle\theta, x_s\rangle - y_s|^2] \leq \sigma_{noise}^2$. For $s = 1, \ldots, \kappa$, assume that $\mathbb{E}[|\langle\theta, x_s\rangle - y_s|^2] \leq \sigma_{noise}^2$. Let $P \in [0,1]$ be arbitrary and let $Q_s(x_s) \sim$ Bernoulli(P) independently for all $s \in [n]$. Let $\Sigma_P = \mathbb{E}_{X\sim\nu}[P(X)XX^\top]$ and $\widehat{\Sigma}_P = \frac{1}{\kappa}\sum_{s=1}^{\kappa} P(\widetilde{x}_s)\widetilde{x}_s\widetilde{x}_s^\top$. Assume that Σ_P is invertible and that there exists $\gamma \geq 0$ such that $(1-\gamma)\Sigma_P \preceq \widehat{\Sigma}_P \preceq (1+\gamma)\Sigma_P$. For a given finite set $\mathcal{V} \subset \mathbb{R}^d$ define*

$$w_v = \mathrm{Catoni}(\{\langle v, \widehat{\Sigma}_P^{-1}Q_s(x_s)x_s y_s\rangle\}_{s=1}^n),$$

If $\widehat{\theta} = \arg\min_\theta \max_v \frac{|w_v - \langle\theta, v\rangle|}{\|v\|_{\widehat{\Sigma}_P^{-1}}}$ and $n \geq 4\log(2|\mathcal{V}|/\delta)$, then with probability at least $1-\delta$, it holds that

$$|\langle v, \widehat{\theta} - \theta\rangle| \leq 4\left(\sqrt{\frac{B^2+\sigma^2}{(1-\gamma)^2}} + \sqrt{n\gamma}\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]}\right)\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}\sqrt{\log(2\,|\mathcal{V}|\,/\delta)}$$

We first state an intermediate matrix lemma before the proof of Proposition 5.

Lemma 53. *Assume that Σ_P is invertible and that there exists $\gamma \in [0, 1/2]$ such that $(1-\gamma)\Sigma_P \preceq \widehat{\Sigma}_P \preceq (1+\gamma)\Sigma_P$. Then for any $v \in \mathcal{V}$*

$$\|v\|^2_{\widehat{\Sigma}_P^{-1}\Sigma_P\widehat{\Sigma}_P^{-1}} \leq \frac{1}{(1-\gamma)^2}\|v\|^2_{\Sigma_P^{-1}}.$$

and

$$\|v\|_{(I-\Sigma_P^{1/2}\widehat{\Sigma}_P^{-1}\Sigma_P^{1/2})^2} \leq \sqrt{1 - \frac{2}{1+\gamma} + \frac{1}{(1-\gamma)^2}}\|v\|_2 \leq \sqrt{10\gamma}\|v\|_2.$$

Proof. We know that taking the inverse of two ordered positive definite matrices will flip the order, so here

$$\frac{1}{(1+\gamma)}\Sigma_P^{-1} \preceq \widehat{\Sigma}_P^{-1} \preceq \frac{1}{(1-\gamma)}\Sigma_P^{-1}.$$

$(1-\gamma)\Sigma_P \preceq \widehat{\Sigma}_P$ implies that for all $u \in \mathbb{R}^d$ holds $u^\top \Sigma_P u \leq 1/(1-\gamma) u^\top \widehat{\Sigma}_P u$. So taking $u = \widehat{\Sigma}_P^{-1} v$, we get $v^\top \widehat{\Sigma}_P^{-1}\Sigma_P \widehat{\Sigma}_P^{-1} v \leq 1/(1-\gamma) v^\top \widehat{\Sigma}_P^{-1} v$. Conclusion

$$v^\top \widehat{\Sigma}_P^{-1}\Sigma_P \widehat{\Sigma}_P^{-1} v = \frac{1}{1-\gamma} v^\top \widehat{\Sigma}_P^{-1} v \leq \frac{1}{(1-\gamma)^2} v^\top \Sigma_P^{-1} v$$

hence the first result of Lemma 53.
For the second one, we get

$$\begin{aligned}
\|v\|^2_{\left(I-\Sigma_P^{1/2}\hat{\Sigma}_P^{-1}\Sigma_P^{1/2}\right)^2} &= v^\top \left(I - \Sigma_P^{1/2}\hat{\Sigma}_P^{-1}\Sigma_P^{1/2}\right)^2 v \\
&= \|v\|_2^2 - 2v^\top \Sigma_P^{1/2}\hat{\Sigma}_P^{-1}\Sigma_P^{1/2} v + v^\top \Sigma_P^{1/2}\hat{\Sigma}_P^{-1}\Sigma_P\hat{\Sigma}_P^{-1}\Sigma_P^{1/2} v \\
&\overset{\text{(i)}}{\leq} \|v\|_2^2 - \frac{2}{1+\gamma}\|v\|_2^2 + \frac{1}{1-\gamma} v^\top \Sigma_P^{1/2}\hat{\Sigma}_P^{-1}\Sigma_P^{1/2} v \\
&\leq \|v\|_2^2 - \frac{2}{1+\gamma}\|v\|_2^2 + \frac{1}{(1-\gamma)^2}\|v\|_2^2 \qquad \text{(Since } \hat{\Sigma}_P \preceq \tfrac{1}{1-\gamma}\Sigma_P\text{)} \\
&\leq \left(1 - \frac{2}{1+\gamma} + \frac{1}{(1-\gamma)^2}\right)\|v\|_2^2 \\
&\overset{\text{(ii)}}{\leq} 10\gamma \|v\|_2^2 .
\end{aligned}$$

The inequality (i) above holds because $\frac{1}{1+\gamma}\Sigma_P^{-1} \preceq \hat{\Sigma}_P^{-1}$ and $(1-\gamma)\Sigma_P \preceq \hat{\Sigma}_P \implies \Sigma_P \preceq \frac{1}{1-\gamma}\hat{\Sigma}_P$. The inequality (ii) above holds because for $\gamma \in \left[0, \frac{1}{2}\right]$, we have

$$1 - \frac{2}{1+\gamma} + \frac{1}{(1-\gamma)^2} \leq 1 - 2(1-\gamma) + (1+2\gamma)^2 \leq 10\gamma.$$

Taking square root on both sides gives us the results. □

Proof of Proposition 5. This proof is analogous to the proof of Proposition 3. We first note that

$$\begin{aligned}
\max_{v\in\mathcal{V}} \frac{|\langle\widehat{\theta}, v\rangle - \langle\theta, v\rangle|}{\|v\|_{\widehat{\Sigma}_P^{-1}}} &= \max_{v\in\mathcal{V}} \frac{|\langle\widehat{\theta}, v\rangle - w_v + w_v - \langle\theta, v\rangle|}{\|v\|_{\widehat{\Sigma}_P^{-1}}} \\
&\leq \max_{v\in\mathcal{V}} \frac{|\langle\widehat{\theta}, v\rangle - w_v|}{\|v\|_{\widehat{\Sigma}_P^{-1}}} + \max_{v\in\mathcal{V}} \frac{|w_v - \langle\theta, v\rangle|}{\|v\|_{\widehat{\Sigma}_P^{-1}}} \\
&= \min_{\theta'} \max_{v\in\mathcal{V}} \frac{|\langle\theta', v\rangle - w_v|}{\|v\|_{\widehat{\Sigma}_P^{-1}}} + \max_{v\in\mathcal{V}} \frac{|w_v - \langle\theta', v\rangle|}{\|v\|_{\widehat{\Sigma}_P^{-1}}} \\
&\leq 2\max_{v\in\mathcal{V}} \frac{|\langle\theta, v\rangle - w_v|}{\|v\|_{\widehat{\Sigma}_P^{-1}}}
\end{aligned}$$

So it suffices to show that each $|\langle \theta, v\rangle - w_v|$ is small. We begin by fixing some $v \in \mathcal{V}$ and bounding the variance of $v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s$ for any $s \le n$ which is necessary to use the robust estimator. Note that

$$\begin{aligned}\mathbb{V}\mathrm{ar}_{x_s\sim\nu, Q_s(x_s)\sim P(x_s)}(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s) =& \mathbb{E}_{x_s\sim\nu, Q_s(x_s)\sim P(x_s)}[(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s)^2] \\ &- \mathbb{E}_{x_s\sim\nu, Q_s(x_s)\sim P(x_s)}[v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s]^2\end{aligned}$$

which means we can drop the second term to bound the variance by

$$\begin{aligned}&\mathbb{E}_{x_s\sim\nu, Q_s(x_s)\sim P(x_s)}[\Big((v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s\Big)^2] \\ &= \mathbb{E}_{x_s\sim\nu, Q_s(x_s)\sim P(x_s)}[\Big(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s (x_s^\top \theta + \xi_s)\Big)^2] \\ &= \mathbb{E}_{x_s\sim\nu}\left[\mathbb{E}_{Q_s(x_s)\sim P(s_s)}[\Big(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s (x_s^\top \theta)\Big)^2] + \mathbb{E}_{Q_s(x_s)\sim P(s_s)}[\Big(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s\Big)^2 \xi_t^2]\right] \\ &\le \mathbb{E}_{x_s\sim\nu}\left[B^2 \mathbb{E}_{Q_s(x_s)\sim P(s_s)}[\Big(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s\Big)^2] + \sigma^2 \mathbb{E}_{Q_s(x_s)\sim P(s_s)}[\Big(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s\Big)^2]\right] \\ &= \mathbb{E}_{x_s\sim\nu}\left[(B^2+\sigma^2)\mathbb{E}_{Q_s(x_s)\sim P(s_s)}[v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s x_s^\top Q_s(x_s) \widehat{\Sigma}_P^{-1} v]\right] \\ &= \mathbb{E}_{x_s\sim\nu}\left[(B^2+\sigma^2)\mathbb{E}_{Q_s(x_s)\sim P(s_s)}[v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s x_s^\top \widehat{\Sigma}_P^{-1} v]\right] \\ &\le \mathbb{E}_{x_s\sim\nu}\left[(B^2+\sigma^2) v^\top \widehat{\Sigma}_P^{-1} P(x_s) x_s x_s^\top \widehat{\Sigma}_P^{-1} v\right],\end{aligned}$$

where we used that $Q_s^2(x_s) = Q_s(x_s)$. Thus, we have with Lemma 53

$$\begin{aligned}\mathbb{V}\mathrm{ar}(v^\top \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s) &\le (B^2+\sigma^2) v^\top \widehat{\Sigma}_P^{-1} \mathbb{E}_{x_s\sim\nu}[P(x_s) x_s x_s^\top] \widehat{\Sigma}_P^{-1} v \\ &= (B^2+\sigma^2)\|v\|^2_{\widehat{\Sigma}_P^{-1}\Sigma_P\widehat{\Sigma}_P^{-1}} \\ &\le \frac{B^2+\sigma^2}{(1-\gamma)^2}\|v\|^2_{\Sigma_P^{-1}}.\end{aligned}$$

We have

$$\begin{aligned}|\langle \theta_*, v\rangle - w_v| &= |\langle \theta_*, v\rangle - \mathbb{E}[v^\top \widehat{\Sigma}_P^{-1} P(x_1) x_1 y_1] + \mathbb{E}[v^\top \widehat{\Sigma}_P^{-1} P(x_1) x_1 y_1] - w_v| \\ &\le |\langle \theta_*, v\rangle - \mathbb{E}[v^\top \widehat{\Sigma}_P^{-1} P(x_1) x_1 y_1]| \\ &\quad + |\mathrm{Catoni}(\{\langle v, \widehat{\Sigma}_P^{-1} Q_s(x_s) x_s y_s\rangle\}_{s=1}^n) - \mathbb{E}_{X\sim\nu}[v^\top \widehat{\Sigma}_P^{-1} P(X) X Y]|.\end{aligned}$$

We now recall that we can write $y_t = x_t^\top \theta_* + \xi_t$ where ξ_t is a mean-zero, independent random variable with variance at most σ^2. Thus, using Cauchy-Schwarz and applying Lemma 53, we get

$$\begin{aligned}|\langle \theta_*, v\rangle - \mathbb{E}[v^\top \widehat{\Sigma}_P^{-1} P(x_1) x_1 y_1]| &= |v^\top \theta_* - v^\top \widehat{\Sigma}_P^{-1} \Sigma_P \theta_*| \\ &= |v^\top (I - \widehat{\Sigma}_P^{-1}\Sigma_P)\theta_*| \\ &= |v^\top \Sigma_P^{-1/2}(I - \Sigma_P^{1/2}\widehat{\Sigma}_P^{-1}\Sigma_P^{1/2})\Sigma_P^{1/2}\theta_*| \\ &\le \|\Sigma_P^{-1/2} v\| \, \|\Sigma_P^{1/2}\theta_*\|_{(I-\Sigma_P^{1/2}\widehat{\Sigma}_P^{-1}\Sigma_P^{1/2})^2} \\ &\le \sqrt{10\gamma}\|\Sigma_P^{-1/2} v\| \, \|\Sigma_P^{1/2}\theta_*\|\end{aligned}$$

$$= \sqrt{10\gamma}\|v\|_{\Sigma_P^{-1}}\|\theta_*\|_{\Sigma_P}.$$

By using the property of Catoni estimator stated in Definition 7, we have

$$\begin{aligned}
&|\langle \theta_*, v\rangle - w_v| \\
\leq& |\text{Catoni}(\{\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q_s(x_s)x_sy_s\rangle\}_{s=1}^n) - \mathbb{E}[\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q_s(x_s)x_sy_s\rangle]| \\
&\quad + \sqrt{10\gamma}\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]}\|v\|_{(\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}} \\
\leq& \sqrt{2}\sqrt{(\mathbb{V}\text{ar}(\langle v, \mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}Q_s(x_s)x_sy_s\rangle))\frac{\log(\frac{2}{\delta})}{n/2}} \\
&\quad + \sqrt{10\gamma}\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]}\|v\|_{(\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(with probability at least } 1-\delta \text{ if } n \geq 4\log(2/\delta)) \\
\leq& \left(\sqrt{4}\sqrt{\frac{B^2+\sigma^2}{(1-\gamma)^2}} + \sqrt{10n\gamma}\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]}\right)\|v\|_{(\mathbb{E}_{X\sim\nu}[P(X)XX^\top]^{-1}}\sqrt{\frac{\log(\frac{2}{\delta})}{n}} \\
=& \left(\sqrt{4}\sqrt{\frac{B^2+\sigma^2}{(1-\gamma)^2}} + \sqrt{10n\gamma}\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]}\right)\|v\|_{\mathbb{E}_{X\sim\nu}[nP(X)XX^\top]^{-1}}\sqrt{\log(2/\delta)}.
\end{aligned}$$

Finally, the proof is complete by taking union bounding over all $v \in \mathcal{V}$. □

Proof of Lemma 51. Most of this proof is exactly the one of Section C.2.1 and Section C.2.1 so we only state the concentration bound. For any $\mathcal{V} \subseteq \mathcal{Z}$ and $z, z' \in \mathcal{V}$ define

$$\mathcal{E}_{z,z',\ell}(\mathcal{V}) = \{|\langle z - z', \widehat{\theta}_\ell(\mathcal{V}) - \theta_*\rangle| \leq \epsilon_\ell\}$$

where $\widehat{\theta}_\ell(\mathcal{V})$ is the estimator that would be constructed by the algorithm at stage ℓ with $\mathcal{Z}_\ell = \mathcal{V}$. Naturally we want to apply Proposition 5 with τ labeled samples to obtain that $\mathcal{E}_{z,z',\ell}(\mathcal{V})$ holds with probability at least $1 - \frac{\delta}{2\ell^2|\mathcal{Z}|^2}$. Note that as Lemma 48 gives $P(x) \geq \mu/3$ so

$$\Sigma_P = \mathbb{E}_{X\sim\nu}[P(X)XX^\top] \geq \frac{\mu}{3}\mathbb{E}_{X\sim\nu}[XX^\top]$$

Σ_P is invertible.

Defining $\delta_0 := \frac{\delta}{4\ell^2|\mathcal{Z}|^2}$ and setting $\kappa \geq 2c_{\delta_0}\max\{1, 20\|\theta_*\|^2_{\mathbb{E}_{X\sim\nu}[XX^\top]}\}$ where we recall that was defined $c_\delta = K_{\psi_2}^2(\sqrt{d\ln 9/c_1} + \sqrt{\frac{\log(2/\delta)}{c_1}})$, Lemma 52 leads to

$$\frac{c_{\delta_0}}{\kappa} \leq \frac{1}{2}\min\left\{1, \frac{1}{20\|\theta_*\|^2_{\mathbb{E}_{X\sim\nu}[XX^\top]}}\right\}$$

so that we can set $\gamma = c_{\delta_0}/(\tau\kappa)$ in the bound of Proposition 5 to get

$$\sqrt{10\tau\gamma}\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[XX^\top]} \leq \frac{1}{2}$$

and

$$\sqrt{\frac{B^2+\sigma^2}{(1-\gamma)^2}} \leq 2\sqrt{B^2+\sigma^2}$$

So for $\delta_0 = \frac{\delta}{4\ell^2|\mathcal{Z}|^2}$ the event $\widetilde{\mathcal{E}}_{\text{cov}}$ defined as

$$\widetilde{\mathcal{E}}_{\text{cov}} := \left\{ \left(1 - \frac{c_{\delta_0}}{\sqrt{\kappa}}\right) x^\top \Sigma_P x \leq x^\top \widehat{\Sigma}_P x \leq \left(1 + \frac{c_{\delta_0}}{\sqrt{\kappa}}\right) x^\top \Sigma_P x \right\}.$$

happen with probability at least $1 - \delta_0$.

Now, let us for now condition on $\widetilde{\mathcal{E}}_{\text{cov}}$. For fixed $\mathcal{V} \subset \mathcal{Z}$ and $\ell \in \mathbb{N}$ we apply Proposition 5, instantiating the arbitrary P to $\widehat{P}_\ell$ (obtained with OPTIMIZEDESIGN, recall Section C.4.1) so that with probability at least $1 - \frac{\delta}{4\ell^2|\mathcal{Z}|^2}$ we have that for any $z, z' \in \mathcal{V}$ holds that the event $\widetilde{\mathcal{E}}_{\text{RIPS},z,z'}$ defined as

$$\begin{aligned}\widetilde{\mathcal{E}}_{\text{RIPS},z,z'} := \Bigg\{ & |\langle z - z', \widehat{\theta}_\ell(\mathcal{V}) - \theta_* \rangle| \\ & \leq 2\|z - z'\|_{\mathbb{E}_{X\sim\nu}[\tau \widehat{P}_\ell(X) X X^\top]^{-1}} \left(4\sqrt{B^2 + \sigma^2} + 1\right) \sqrt{\log(4\ell^2|\mathcal{Z}|^2/\delta)} \Bigg\}\end{aligned}$$

happen with probability at least $1 - \delta_0$.

So with probability at least $1 - \mathbb{P}(\widetilde{\mathcal{E}}^c_{\text{RIPS},z,z'}) - \mathbb{P}(\widetilde{\mathcal{E}}^c_{\text{cov}}) \geq 1 - \frac{\delta}{4\ell^2|\mathcal{Z}|^2} - \frac{\delta}{4\ell^2|\mathcal{Z}|^2} = 1 - \frac{\delta}{2\ell^2|\mathcal{Z}|^2}$, both events hold and we have that for any $z, z' \in \mathcal{V}$ holds

$$\begin{aligned}|\langle z - z', \widehat{\theta}_\ell(\mathcal{V}) - \theta_* \rangle| &\leq 2\|z - z'\|_{\mathbb{E}_{X\sim\nu}[\tau \widehat{P}_\ell(X) X X^\top]^{-1}} \left(4\sqrt{B^2 + \sigma^2} + 1\right) \sqrt{\log(4\ell^2|\mathcal{Z}|^2/\delta)} \\ &\leq 2(1+\varepsilon) \left(4\sqrt{B^2 + \sigma^2} + 1\right) \|z - z'\|_{\mathbb{E}_{X\sim\nu}[\tau \widehat{P}_\ell(X) X X^\top]^{-1}} \sqrt{\log(4\ell^2|\mathcal{Z}|^2/\delta)} \\ &\leq \epsilon_\ell.\end{aligned}$$

where we used the property of $\widehat{P}_\ell$ as detailed in Section C.4.1 to conclude. □

Proof of Theorem 25. The total number of labels requested after L rounds is equal to $\sum_{\ell=1}^{L} \sum_{t=(\ell-1)\tau+1}^{\ell\tau} \widehat{P}_\ell(x_t)$. Again by Freedman's inequality we have that

$$\sum_{\ell=1}^{L} \sum_{t=(\ell-1)\tau+1}^{\ell\tau} \widehat{P}_\ell(x_t) \leq 2 \sum_{\ell=1}^{L} \tau \mathbb{E}_{X\sim\nu}[\widehat{P}_\ell(X)|\mathcal{Z}_\ell] + \log(1/\delta)$$

From Theorem 13, it holds for any ℓ that $\mathbb{E}_{X\sim\nu}[\widehat{P}_\ell(X)] \leq \mathbb{E}_{X\sim\nu}[\widetilde{P}_\ell(X)] + 4\sqrt{\mu}$ where $\widetilde{P}_\ell$ is the optimal solution to problem (C.13). So now, for some $\widetilde{\tau}$, we want to relate $\mathbb{E}_{X\sim\nu}[\widetilde{\tau}\widetilde{P}_\ell(X)]$ to $\mathbb{E}_{X\sim\nu}[\tau P_\ell(X)]$ where P_ℓ is the solution of problem (4.4). To do so, we rewrite problem (4.4) and problem (C.13) as

$$\begin{aligned} \min_P \quad & \mathbb{E}_{X\sim\nu}\left[\tau P(X)\right] \\ \text{subject to} \quad & y^\top \mathbb{E}_{X\sim\nu}\left[\tau P(X) X X^\top\right]^{-1} y \leq c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell, \\ & 0 \leq \tau P(x) \leq \tau, \quad \forall x \in \mathcal{X}. \end{aligned} \tag{C.25}$$

and

$$\begin{aligned} \min_P \quad & \mathbb{E}_{X\sim\nu}\left[\widetilde{\tau} P(X)\right] \\ \text{subject to} \quad & y^\top \mathbb{E}_{X\sim\nu}[\widetilde{\tau} P(X) X X^\top]^{-1} y \leq c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell, \\ & 0 \leq \widetilde{\tau} P(x) \leq \widetilde{\tau}(1 - \mu_b), \quad \forall x \in \mathcal{X}. \end{aligned} \tag{C.26}$$

where problem (C.25) is equivalent to problem (4.4) and problem (C.26) is equivalent to problem (C.13). Thus taking $\widetilde{\tau} = \frac{\tau}{1-\mu_b}$, problem (C.26) becomes

$$\begin{array}{rl} \min_P & \mathbb{E}_{X\sim\nu}\left[\frac{\tau}{1-\mu_b}P(X)\right] \\ \text{subject to} & y^\top \mathbb{E}_{X\sim\nu}[\frac{\tau}{1-\mu_b}P(X)XX^\top]^{-1}y \le c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell, \\ & 0 \le \frac{\tau}{1-\mu_b}P(x) \le \tau, \quad \forall x \in \mathcal{X}. \end{array}$$

which, using $Q = \frac{P}{1-\mu_b}$ is equivalent to

$$\begin{array}{rl} \min_Q & \mathbb{E}_{X\sim\nu}\left[\tau Q(X)\right] \\ \text{subject to} & y^\top \mathbb{E}_{X\sim\nu}[\tau Q(X)XX^\top]^{-1}y \le c_\ell^2, \quad \forall y \in \mathcal{Y}_\ell, \\ & 0 \le \tau Q(x) \le \tau, \quad \forall x \in \mathcal{X}. \end{array} \tag{C.27}$$

And we can now see that (C.27) and (C.25) are the same optimization problem. And Q_ℓ^* the solution of (C.27) is equal to $\frac{\widetilde{P}_\ell}{1-\mu_b}$. Thus the result $\mathbb{E}_{X\sim\nu}\left[\tilde{\tau}\widetilde{P}_\ell(X)\right] = \mathbb{E}_{X\sim\nu}\left[\tau P_\ell(X)\right]$.

Remains to bound $\sum_{\ell=1}^L \tau\mathbb{E}_{X\sim\nu}[P_\ell(X)]$ where

$$\begin{aligned} &\sum_{\ell=1}^L \tau\mathbb{E}_{X\sim\nu}[P_\ell(X)|\mathcal{Z}_\ell] \\ &= \sum_{\ell=1}^L \left[\min_{P:\mathcal{X}\to[0,1]} \tau\mathbb{E}_{X\sim\nu}[P(X)] \quad \text{subject to} \quad \max_{z,z'\in\mathcal{Z}_\ell} \frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\epsilon_\ell^2}\beta_{\delta,\ell} \le 1\right], \end{aligned}$$

where $\beta_{\delta,\ell}$ is defined in Section C.4.1 as

$$\beta_{\delta,\ell} := 4(1+\varepsilon)^2\left(4\sqrt{B^2+\sigma^2}+1\right)^2\log(4\ell^2|\mathcal{Z}|^2/\delta).$$

As in the case where the distribution ν is known (Section C.2.1), we use Lemma 37 to bound $\max_{z,z'\in\mathcal{Z}_\ell} \frac{\|z-z'\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top}}{\epsilon_\ell^2}$ by $\max_{z\in\mathcal{Z}\setminus z_*} \frac{\|z-z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X)XX^\top]^{-1}}}{\langle z-z_*,\theta_*\rangle^2} 64\beta_{\delta,L}$. Last, the reparameterization of Proposition 4 also applies here.

In the unlabeled sample complexity, we get an additional $L\kappa = L\lceil 2K_{\psi_2}^2(\sqrt{d\ln 9/c_1}+\sqrt{\frac{\log(2/\delta)}{c_1}})\max\{1, 20\|\theta_*\|_{\mathbb{E}_{X\sim\nu}[X}$ term from the estimation of the covariance matrix. Last, we get an additional $L(K+u)$, where K and u are such that

$$K \ge \widetilde{O}\left(\frac{|\mathcal{Z}|^3\,\kappa(\Sigma)^2\,\|\Lambda^*\|_2^8\,M^{16}}{\beta^2\mu_b^6}\right)\cdot\left(\frac{1+\epsilon}{\epsilon}\right)^2, \quad u \ge \widetilde{O}\left(\frac{\kappa(\Sigma)^2\,\|\Lambda^*\|_2^6\,M^{16}}{\beta^2\mu_b^6}\right)\cdot\left(\frac{1+\epsilon}{\epsilon}\right)^2,$$

from the sample complexity of the subroutine. □

C.5 Classification

In this section we adopt the implementation described in Section C.2.1. As described in the text, given a distribution $\pi \in \Delta_\mathcal{X}$, and a class of hypothesis $\mathcal{H}$, we can reduce classification to linear bandits by setting $\theta^* = [\theta_x^*]_{x\in\Delta_\mathcal{X}}$ where $\theta_x^* = 2\eta(x) - 1$, and $\mathcal{Z} := \{z^{(h)}\}_{h\in\mathcal{H}} \subset [0,1]^{|\mathcal{X}|}$ where $z_x^{(h)} = \pi(x)\mathbf{1}\{h(x) = 1\}$. With the quantities computed in Section 4.3, we now prove Theorem 12.

Proof of Theorem 12. We consider a slightly modified version of Algorithm 4.1 where we stop at round L where $L_\epsilon = \lceil \log_2(4/\epsilon) \rceil$ and return $\arg\max_{z^{(h)} \in \mathcal{Z}_\ell} \langle z^{(h)}, \widehat{\theta}_\ell \rangle$. By an identical analysis to that in the proof of Theorem 2, we are guaranteed that $h \in \mathcal{S}_\ell$, i.e. $R_\nu(h) - R_\nu(z^*) = \langle z^* - z, \theta_* \rangle \leq 4\epsilon_\ell$. In addition the analysis of the sample complexity given there immediately gives the first part of the theorem.

It remains to bound the sample complexity in terms of the disagreement coefficient. The total sample complexity is given by,

$$\sum_{\ell=1}^{L} \left[\min_{P:\mathcal{X}\to[0,1]} \tau \mathbb{E}_{X\sim\nu}[P(X)] \quad \text{subject to} \quad \max_{z\in\mathcal{S}_\ell} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X) XX^\top]^{-1}}}{\epsilon_\ell^2} \beta_\delta \leq 1 \right]$$

where we recall $\beta_\delta = 2048 \log(2L^2|\mathcal{H}|/\delta)$ since we can take $B = 1$ and $\sigma = 1$.

We recall the proof of Theorem 2. From the proof, we see that with probability greater than $1 - \delta$, our sample complexity is obtained by summing up to round L

$$\sum_{\ell=1}^{L} \left[\min_{P:\mathcal{X}\to[0,1]} \tau \mathbb{E}_{X\sim\nu}[P(X)] \quad \text{subject to} \quad \max_{z\in\mathcal{S}_\ell} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\nu}[\tau P(X) XX^\top]^{-1}}}{\epsilon_\ell^2} \beta_\delta \leq 1 \right]$$

By proposition 2 this is equivalent to

$$\sum_{\ell=1}^{L} \left[\min_{\lambda\in\Delta_X} \rho_\ell(\lambda)\beta_\delta \quad \text{subject to} \quad \left\| \frac{\lambda}{\nu} \right\|_\infty \rho_\ell(\lambda)\beta_\delta \leq \tau \right], \text{ where } \rho_\ell(\lambda) := \max_{z\in\mathcal{S}_\ell} \frac{\|z - z_*\|^2_{\mathbb{E}_{X\sim\lambda}[XX^\top]^{-1}}}{\epsilon_\ell^2}.$$

Define

$$A_\ell = \{x \in \mathcal{X} : \exists h, h(x) \neq h^*(x), R_\nu(h) - R_\nu(h^*) \leq 4\epsilon_\ell\}, \ell \leq L$$

and let $\lambda_\ell = \dfrac{\mathbf{1}\{x \in A_\ell\}\nu(x)}{\mathbb{E}[\mathbf{1}\{x \in A_\ell\}]}$, so $\left\| \dfrac{\lambda}{\nu} \right\|_\infty = \dfrac{1}{\mathbb{E}[\mathbf{1}\{x \in A_i\}]}$.

We first argue that λ_ℓ is feasible for the previous program. Note,

$$\begin{aligned}
\rho_\ell(\lambda_\ell) &= \max_{h:R_\nu(h)-R_\nu(h^*)\leq 4\epsilon_\ell} \frac{\mathbb{E}_{X\sim\nu}[\frac{\mathbf{1}\{h(x)\neq h^*(x)\}}{\lambda_\ell(x)/\nu(x)}\}]}{\epsilon_\ell^2} \\
&\overset{\text{(i)}}{=} \mathbb{E}[\mathbf{1}\{x \in A_\ell\}] \max_{h:R_\nu(h)-R_\nu(h^*)\leq 4\epsilon_\ell} \frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(x) \neq h^*(x)\}]}{\epsilon_\ell^2} \\
&\leq \mathbb{E}[\mathbf{1}\{x \in A_\ell\}] \max_{h:R_\nu(h)-R_\nu(h^*)\leq 4\epsilon_\ell} \frac{16\mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(x) \neq h^*(x)\}]}{\max\{\epsilon_\ell^2, (R_\nu(h) - R_\nu(h^*))^2\}} \\
&\leq \mathbb{E}[\mathbf{1}\{x \in A_\ell\}] \max_{h:R_\nu(h)-R_\nu(h^*)\leq 4\epsilon_\ell} \frac{16\mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(x) \neq h^*(x)\}]}{\max\{(4\epsilon_\ell)^2, (R_\nu(h) - R_\nu(h^*))^2\}} \\
&\overset{\text{(ii)}}{\leq} \mathbb{E}[\mathbf{1}\{x \in A_\ell\}] \max_{h:R_\nu(h)-R_\nu(h^*)\leq 4\epsilon_\ell} \frac{16\mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(x) \neq h^*(x)\}]}{\max\{\epsilon^2, (R_\nu(h) - R_\nu(h^*))^2\}} \\
&\leq \mathbb{E}[\mathbf{1}\{x \in A_\ell\}] \max_{h\in H} \frac{16\mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(x) \neq h^*(x)\}]}{\max\{\epsilon^2, (R_\nu(h) - R_\nu(h^*))^2\}} \\
&\leq 16\mathbb{E}[\mathbf{1}\{x \in A_\ell\}]\rho(\nu, \epsilon)
\end{aligned}$$

where the equality (i) holds because the following is true when we only consider h such that $R_\nu(h) - R_\nu(h^*) \leq 4\epsilon_\ell$

$$\frac{\mathbf{1}\{h(x) \neq h^*(x)\}}{\mathbf{1}\{x : \exists h, h(x) \neq h^*(x), (R_\nu(h) - R_\nu(h^*)) \leq 4\epsilon_\ell\}} = \mathbf{1}\{h(x) \neq h^*(x)\}.$$

The inequality (ii) above is true because $4\epsilon_\ell \geq \epsilon$. Thus we see that $\rho_\ell(\lambda_\ell)\|\lambda/\nu\|_\infty \beta_\delta \leq 16\rho(\nu,\epsilon)\beta_\delta \leq \tau$. It remains to argue about the disagreement coefficient. Firstly note that for any h such that $R_\nu(h) - R_\nu(h^*) \leq 4\epsilon_\ell$.

$$
\begin{aligned}
d_\nu(h,h^*) = \mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(X) \neq h^*(X)\}] &\leq \mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(X)\neq Y\}] + \mathbb{E}_{X\sim\nu}[\mathbf{1}\{h^*(X)\neq Y\}] && \text{(C.28)}\\
&\leq R_\nu(h) + R_\nu(h^*) && \text{(C.29)}\\
&\leq 2R_\nu(h^*) + 4\epsilon_\ell && \text{(C.30)}
\end{aligned}
$$

Using this we see that,

$$
\begin{aligned}
&\min_{\lambda\in\Delta} \rho_\ell(\lambda) \text{ subject to } \rho_\ell(\lambda)\|\lambda/\nu\|_\infty\beta_\delta \leq \tau\\
&\qquad \leq \rho_\ell(\lambda_\ell)\beta_\delta && \text{(since } \lambda_\ell \text{ is feasible.)}\\
&\qquad \leq \mathbb{E}[\mathbf{1}\{x\in A_\ell\}] \max_{h:R_\nu(h)-R_\nu(h^*)\leq 4\epsilon_\ell} \frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{h(x)\neq h^*(x)\}]}{\epsilon_\ell^2}\beta_\delta\\
&\qquad && \text{(imitating the above computation)}\\
&\qquad \leq \frac{(2R(h^*)+4\epsilon_\ell)\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+4\epsilon_\ell\}]}{\epsilon_\ell^2}\beta_\delta\\
&\qquad && \text{(Equation (C.28))}\\
&\qquad \leq \beta_\delta \begin{cases} \frac{9R(h^*)^2}{\epsilon_\ell^2}\frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+4\epsilon_\ell\}]}{2R(h^*)+4\epsilon_\ell} & 4\epsilon_\ell \leq R(h^*)\\ \frac{144\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+4\epsilon_\ell\}]}{2R(h^*)+4\epsilon_\ell} & 4\epsilon_\ell > R(h^*)\end{cases}\\
&\qquad \leq \left(\frac{9R(h^*)^2}{\epsilon_\ell^2}+144\right)\frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+4\epsilon_\ell\}]}{2R(h^*)+4\epsilon_\ell}\beta_\delta
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sum_{\ell=1}^{L}\left[\min_{\lambda\in\Delta_X}\rho_\ell(\lambda)\beta_\delta \quad \text{subject to} \quad \left\|\frac{\lambda}{\nu}\right\|_\infty \rho_\ell(\lambda)\beta_\delta \leq \tau\right]\\
&\leq \sum_{\ell=1}^{L}\rho_\ell(\lambda_\ell)\beta_\delta\\
&\leq \sum_{\ell=1}^{L}\left(\frac{9R(h^*)^2}{\epsilon_\ell^2}+144\right)\frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+4\epsilon_\ell\}]}{2R(h^*)+4\epsilon_\ell}\beta_\delta\\
&\leq \log_2\left(\frac{4}{\epsilon}\right)\sup_{\ell\leq L}\left(\frac{9R(h^*)^2}{\epsilon_\ell^2}+144\right)\frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+\epsilon_\ell\}]}{2R(h^*)+4\epsilon_\ell}\beta_\delta\\
&\leq \log_2\left(\frac{4}{\epsilon}\right)\left(\frac{36R(h^*)^2}{\epsilon^2}+144\right)\sup_{\ell\leq L}\frac{\mathbb{E}_{X\sim\nu}[\mathbf{1}\{\exists h: h(X)\neq h^*(X), d_\nu(h,h^*)\leq 2R(h^*)+4\epsilon_\ell\}]}{2R(h^*)+4\epsilon_\ell}\beta_\delta\\
&\leq 36\log_2\left(\frac{4}{\epsilon}\right)\left(\frac{R(h^*)^2}{\epsilon^2}+4\right)\sup_{\xi\geq\epsilon}\theta^*(2R(h^*)+\xi,\nu)\beta_\delta
\end{aligned}
$$

from which the result follows.

□

Appendix D

Appendix for Chap. 5

D.1 Additional Algorithms in Implementation

D.1.1 A Peace-based Robust Algorithm

In this section, we briefly explain how we design P1-Peace based on intuition similar to P1-RAGE and make it computationally efficient. First, we propose another subroutine, called Peace-Elimination, based on the elimination strategy in Peace Katz-Samuels et al. (2020), which has the same spirit as RAGE. Similar to RAGE-Elimination, Peace-Elimination also repeatedly computes $\mathcal{XY}$-allocation, but (virtually) eliminate arms so that the value of the remaining arms' optimal $\mathcal{XY}$-design is halved. In addition, in P1-Peace, we only update the sampling distribution λ_t after a period of time. The intuition is that if the environment is stationary, then we do not need to update our allocation probability frequently just like RAGE and Peace; if the environment is non-stationary, then the non-stationarity is handled by the mixed G-optimal design λ^*, which is fixed from the very beginning. Therefore, updating λ_t in a low frequency should not severely harm the performance. The new algorithm and elimination subroutine are summarized in Algorithm D.1 and D.2.

For convenience of presentation, for arm set $\mathcal{Z} \subset \mathbb{R}^d$ and distribution $\lambda \in \triangle_{\mathcal{X}}$, we define

$$\rho(\mathcal{Z}, \lambda) = \max_{x, x' \in \mathcal{Z}} \left\| x - x' \right\|^2_{A(\lambda)^{-1}}. \tag{D.1}$$

Algorithm D.1. P1-Peace

Input: budget, $T \in \mathbb{N}$; arm set $\mathcal{X} \subset \mathbb{R}^d$
Compute epoch length $R \leftarrow \left\lfloor \frac{T}{\log_2(\inf_{\lambda \in \triangle_{\mathcal{X}}} \rho(\mathcal{X}, \lambda))} \right\rfloor$
Compute G-optimal design λ^* based on equation (5.3) and initialize $\lambda_1 = \lambda^*$
for $t = 1, 2, \ldots, T$ **do**
    Sample $x_t \sim \lambda_t$ and receive reward r_t
    Estimate $\widehat{\theta}_t \leftarrow \frac{1}{t} \sum_{s=1}^{t} \mathbb{E}_{x \sim \lambda_s} \left[x x^\top \right]^{-1} x_s r_s$
    $\lambda_{t+1} \leftarrow \lambda_t$
    if $t - 1 = cR$ for some integer c **then**
        Update $\lambda_{t+1} \leftarrow$Peace-Elimination$(\widehat{\theta}_t)$
Return $\arg\max_{x \in \mathcal{X}} x^\top \widehat{\theta}_T$

Algorithm D.2. Peace-Elimination

Input: arm set $\mathcal{X} \subset \mathbb{R}^d$; current estimate $\widehat{\theta}_t$
Find index $\widehat{(k)}_t$ such that $x_{\widehat{(1)}_t}^\top \widehat{\theta}_t \geq x_{\widehat{(2)}_t}^\top \widehat{\theta}_t \geq \cdots \geq x_{\widehat{(K)}_t}^\top \widehat{\theta}_t$
Initialize $\mathcal{X}_t^{(0)} \leftarrow \mathcal{X}$ and $i \leftarrow 0$
while $|\mathcal{X}_t^{(i)}| > 1$ **do**
    Compute $\lambda_t^{(i)} \leftarrow \arg\inf_{\lambda \in \triangle_\mathcal{X}} \rho(\mathcal{X}_t^{(i)}, \lambda)$
    Find the largest index k_i such that

$$\inf_{\lambda \in \triangle_\mathcal{X}} \rho\left(\{x_{\widehat{(1)}_t}, \ldots, x_{\widehat{(k_i)}_t}\}\right) \leq \frac{1}{2} \cdot \inf_{\lambda \in \triangle_\mathcal{X}} \rho(\mathcal{X}_t^{(i)}, \lambda)$$

    Update $\mathcal{X}_t^{(i+1)} \leftarrow \left\{x_{\widehat{(1)}_t}, \ldots, x_{\widehat{(k_i)}_t}\right\}$
    $i \leftarrow i + 1$
Return $(\bar{\lambda}_t + \lambda^*)/2$, where $\bar{\lambda}_t = \frac{1}{i}\sum_{i'=0}^{i-1} \lambda_t^{(i')}$

D.1.2 A Naive Baseline Mixed Algorithm

In this section, we present a naive mixture of Peace and the G-optimal design, called Mixed-Peace, which eliminates arms and computes design λ_k during each epoch exactly the same as Peace. The only differences are that Mixed-Peace uses IPS estimator and when pulling an arm, it will pull an arm by following $x_t \sim (\lambda_k + \lambda^*)/2$, where λ^* is the G-optimal design defined in equation (5.3). Its details are summarized in Algorithm D.3.

Algorithm D.3. Mixed-Peace

Input: budget, $T \in \mathbb{N}$; arm set $\mathcal{X} \subset \mathbb{R}^d$
Initialize $R \leftarrow \lceil \log_2 (\inf_{\lambda \in \triangle_\mathcal{X}} \rho(\mathcal{X}, \lambda)) \rceil$, $N \leftarrow \lfloor \frac{T}{R} \rfloor$, $\mathcal{X}_0 \leftarrow \mathcal{X}$, $\widehat{\theta}_0 \leftarrow \mathbf{0}$ and $t \leftarrow 1$
Compute G-optimal design λ^* using equation (5.3)
for $r = 0, \ldots, R$ **do**
    Find $\lambda_r \leftarrow (\arg\inf_{\lambda \in \triangle_\mathcal{X}} \rho(\mathcal{X}_r, \lambda) + \lambda^*)/2$
    while $t \leq \min\{T, (r+1)N\}$ **do**
        Sample $x_t \sim \lambda_r$ and receive reward r_t
        Estimate $\widehat{\theta}_t \leftarrow \frac{t-1}{t} \cdot \widehat{\theta}_{t-1} + \frac{1}{t} \cdot \mathbb{E}_{x \sim \lambda_r}\left[xx^\top\right]^{-1} x_t r_t$
        $t \leftarrow t + 1$
    if $|\mathcal{X}_r| > 1$ **then**
        Reindex $\mathcal{X}_r$ such that $x_1^\top \widehat{\theta}_t \geq x_2^\top \widehat{\theta}_t \geq \cdots \geq x_{n_r}^\top \widehat{\theta}_t$, where $n_r = |\mathcal{X}_r|$
        Find the largest index k_r such that

$$\inf_{\lambda \in \triangle_\mathcal{X}} \rho(\{x_1, \ldots, x_{k_r}\}, \lambda) \leq \frac{1}{2} \cdot \inf_{\lambda \in \triangle_\mathcal{X}} \rho(\mathcal{X}_r, \lambda)$$

        Update $\mathcal{X}_{r+1} \leftarrow \{x_1, \ldots, x_{k_r}\}$
return $\arg\max_{x \in \mathcal{X}} x^\top \widehat{\theta}_T$

D.2 Error Probability of Algorithm 5.1 In Non-Stationary Environments

Recall that Theorem 14 states the following: Fix time horizon T, arm set $\mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = K$ and arbitrary unknown parameters $\{\theta_t\}_{t=1}^T$. If we run Algorithm 5.1 in this non-stationary environment and obtain x_{J_T}, then it holds that

$$\mathbb{P}_{\overline{\theta}_T}\left(J_T \neq (1)\right) \leq K \exp\left(-\frac{T}{12 H_{\text{G-BAI}}\left(\overline{\theta}_T\right)}\right), \quad \text{where } H_{\text{G-BAI}}\left(\overline{\theta}_T\right) = \frac{d}{\Delta_{(1)}^2}.$$

Proof. Based on the recommendation rule $x_{J_T} = \arg\max_{x \in \mathcal{X}} x^\top \widehat{\theta}_T$, we have

$$\begin{aligned}
\mathbb{P}\left(J_T \neq (1)\right) =& \mathbb{P}\left(\exists k \in [2:K] \text{ s.t. } x_{(k)}^\top \widehat{\theta}_T \geq x_{(1)}^\top \widehat{\theta}_T\right) \\
\leq& \mathbb{P}\left(\exists k \in [2:K] \text{ s.t. } x_{(k)}^\top \widehat{\theta}_T - x_{(k)}^\top \overline{\theta}_T \geq \frac{\Delta_{(k)}}{2} \text{ or } x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) \\
\leq& \mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) + \sum_{k=2}^K \mathbb{P}\left(x_{(k)}^\top \widehat{\theta}_T - x_{(k)}^\top \overline{\theta}_T \geq \frac{\Delta_{(k)}}{2}\right). \qquad \text{(D.2)}
\end{aligned}$$

The above terms can be bounded by Bernstein's inequality. In particular, for the first term, we have

$$\mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) = \mathbb{P}\left(\sum_{t=1}^T x_{(1)}^\top \left(A(\lambda^*)^{-1} x_t r_t - \theta_t\right) \leq -\frac{T\Delta_{(1)}}{2}\right).$$

Since IPS estimator is unbiased, $x_{(1)}^\top \left(A(\lambda^*)^{-1} x_t r_t - \theta_t\right)$ is a zero-mean random variable. Based on our bounded reward assumption, we have

$$\left|x_{(1)}^\top \left(A(\lambda^*)^{-1} x_t r_t - \theta_t\right)\right| \leq \left|x_{(1)}^\top A(\lambda^*)^{-1} x_t\right| + 2 \leq \left\|x_{(1)}\right\|_{A(\lambda^*)^{-1}} \left\|x_t\right\|_{A(\lambda^*)^{-1}} + 2 \leq d + 2 \leq 3d,$$

where we use the property of G-optimal design $\max_{x \in \mathcal{X}} \|x\|_{A(\lambda^*)^{-1}}^2 \leq d$. We can similarly bound its variance by

$$\begin{aligned}
\mathbb{E}\left[\left(x_{(1)}^\top \left(A(\lambda^*)^{-1} x_t r_t - \theta_t\right)\right)^2\right] \leq& \mathbb{E}\left[\left(x_{(1)}^\top A(\lambda^*)^{-1} x_t\right)^2\right] \\
=& x_{(1)}^\top A(\lambda^*)^{-1} \mathbb{E}\left[x_t x_t^\top\right] A(\lambda^*)^{-1} x_{(1)} \\
=& x_{(1)}^\top A(\lambda^*)^{-1} A(\lambda^*) A(\lambda^*)^{-1} x_{(1)} \qquad \text{(Since } x_t \sim \lambda^* \text{ by algorithm)} \\
=& \left\|x_{(1)}\right\|_{A(\lambda^*)^{-1}}^2 \leq d
\end{aligned}$$

Thus, by Bernstein's inequality, we have

$$\mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) \leq \exp\left(-\frac{T^2\Delta_{(1)}^2/8}{Td + Td\Delta_{(1)}/2}\right) \leq \exp\left(-\frac{T\Delta_{(1)}^2}{12d}\right),$$

where the last inequality uses the assumption that $\Delta_{(1)} \leq 1$. By similarly applying Bernstein's inequality to other terms in (D.2), we can then have

$$\mathbb{P}\left(J_T \neq x_{(1)}\right) \leq \mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) + \sum_{k=2}^K \mathbb{P}\left(x_{(k)}^\top \widehat{\theta}_T - x_{(k)}^\top \overline{\theta}_T \geq \frac{\Delta_{(k)}}{2}\right)$$

$$\leq \sum_{k=1}^{K} \exp\left(-\frac{T\Delta_{(k)}^2}{12d}\right)$$
$$\leq K \exp\left(-\frac{T\Delta_{(1)}^2}{12d}\right).$$

□

D.3 Error Probability of Algorithm 5.2

D.3.1 Stationary Environments

We first prove an error probability of Algorithm 5.2 in stationary environments that contains unspecified parameters from the virtual phases. Without loss of generality, assume that the arms $x_1, \ldots, x_K$ are ordered such that $\theta^\top x_1 > \theta^\top x_2 \geq \cdots \geq \theta^\top x_K$ and $\Delta_1 = \Delta_2 \leq \Delta_3 \leq \cdots \leq \Delta_K$.

Throughout this section, we will the following definitions: $i_0 = \lceil \log_2(1/\Delta_1) \rceil + 1$, $\mathcal{A}_i = \left\{x \in \mathcal{X} \mid \Delta_x \leq 2 \cdot 2^{-i}\right\}$, $\bar{i}(k) = \max\left\{i \in [i_0 - 1] \mid \Delta_k \leq 2^{-i}\right\}$ and

$$f(\mathcal{A}_i) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x, x' \in \mathcal{A}_i} \left\|x - x'\right\|^2_{A(\lambda)^{-1}}.$$

Theorem 26. *Let $\mathcal{D} = \left\{\mathbf{a} \in [0,1]^{i_0+1} \mid 0 = a_0 < a_1 \leq a_2 \leq \ldots \leq a_{i_0} = 1\right\}$. Then, if $m \geq i_0$, The error probability of Algorithm 5.2 in a stationary environment with parameter θ is bounded as*

$$\mathbb{P}_\theta\left(J_T \neq 1\right) \leq 2 i_0 K T \exp\left(-\frac{T}{\overline{H}_{\textit{P1-RAGE}}(\theta)}\right),$$
$$\overline{H}_{\textit{P1-RAGE}}(\theta) = \min_{\mathbf{a} \in \mathcal{D}} \max_{k \in [K]} \frac{48m \sum_{i'=1}^{\bar{i}(k)} (a_{i'} - a_{i'-1}) f(\mathcal{A}_{i'-2}) + 8(m\sqrt{df(\mathcal{X})} + 1) a_{\bar{i}(k)} \Delta_k}{3 a_{\bar{i}(k)}^2 \Delta_k^2}. \tag{D.3}$$

Proof. With $0 = n_0 < n_1 \leq n_2 \leq \ldots \leq n_{i_0} = T$.[1] we define the event ξ_i with $i \geq 1$ as follows: after n_i samples all the arms with true gap smaller than $2 \cdot 2^{-i}$ are estimated with precision $2^{-i}/2$, which is

$$\xi_i = \{\forall t \geq n_i, \forall k \in [K] \text{ s.t. } \Delta_k \leq 2 \cdot 2^{-i} \implies |\Delta_k - \widehat{\Delta}_k^{(t)}| < 2^{-i}/2\},$$

where $\widehat{\Delta}_k^{(t)} = (x_1 - x_k)^\top \widehat{\theta}^{(t)}$ for $k > 1$ and $\widehat{\Delta}_1^{(t)} = (x_1 - x_2)^\top \widehat{\theta}^{(t)}$. We first show how these events $\{\xi_i\}_{i=1}^{i_0}$ relate the correctness of Algorithm 5.2.

Correctness. If $\bigcap_{i=1}^{i_0} \xi_i$ holds then the algorithm successfully identifies the best arm. Indeed, if we assume it does not, then there must exist non-optimal arm k_0 such that $\widehat{\Delta}_{k_0}^{(T)} < 0$. As $\bigcap_{i=1}^{i_0} \xi_i$ holds, for some $i' \leq i_0$, it holds that $2^{-i'} < \Delta_{k_0} \leq 2 \cdot 2^{-i'}$ and then $|\Delta_{k_0} - \widehat{\Delta}_{k_0}^{(T)}| < 2^{-i'}/2$. Therefore, we have $2^{-i'} < \Delta_{k_0} \leq \Delta_{k_0} - \widehat{\Delta}_{k_0}^{(T)} \leq |\Delta_{k_0} - \widehat{\Delta}_{k_0}^{(T)}| \leq 2^{-i'}/2$, which is a contradiction.

Thus, the error probability is upper bounded by $\mathbb{P}\left(\bigcup_{i=1}^{i_0} \xi_i^c\right)$, which gives us

$$\mathbb{P}\left(J_T \neq 1\right) \leq \mathbb{P}\left(\bigcup_{i=1}^{i_0} \xi_i^c\right) = \mathbb{P}\left(\bigcup_{i=1}^{i_0} \left(\xi_i^c \setminus \bigcup_{j=1}^{i-1} \xi_j^c\right)\right) \leq \sum_{i=1}^{i_0} \mathbb{P}\left(\xi_i^c \setminus \bigcup_{j=1}^{i-1} \xi_j^c\right)$$

[1] We do not specify the values of $n_1, \ldots, n_{i_0 - 1}$ for now.

$$
\begin{aligned}
&= \sum_{i=1}^{i_0} \mathbb{P}\left(\xi_i^c \cap \left(\bigcup_{j=1}^{i-1} \xi_j^c\right)^c\right) = \sum_{i=1}^{i_0} \mathbb{P}\left(\xi_i^c \cap \left(\bigcap_{j=1}^{i-1} \xi_j\right)\right) \\
&\leq \sum_{i=1}^{i_0} \mathbb{P}\left(\xi_i^c \,\middle|\, \bigcap_{j=1}^{i-1} \xi_j\right).
\end{aligned}
$$

Bernstein's inequality. Now, we just need to find an upper bound of $\mathbb{P}\left(\xi_i^c \,\middle|\, \bigcap_{j=1}^{i-1} \xi_j\right)$. Assume $\exists t \geq n_i, \exists k \in [K]$ s.t. $\Delta_k \leq 2 \cdot 2^{-i}$.[2] Then, we have

$$
\begin{aligned}
&\mathbb{P}(|\Delta_k - \widehat{\Delta}_k^{(t)}| \geq 2^{-i}/2) \\
=&\mathbb{P}(|(\theta - \widehat{\theta}_t)^\top (x_1 - x_k)| \geq 2^{-i}/2) \\
=&\mathbb{P}\left(\left|\sum_{s=1}^{t} \left(\theta - A(\lambda_s)^{-1} x_s r_s\right)^\top (x_1 - x_k)\right| \geq 2^{-i} t/2\right) \\
\overset{\text{(a)}}{\leq}& 2\exp\left(-\frac{2^{-2i} t^2/8}{2\sum_{s=1}^{t} \|x_1 - x_k\|^2_{A(\bar{\lambda}_s)^{-1}} + \left(\sqrt{d}\max_{s\in[1:t]} \|x_1 - x_k\|_{A(\bar{\lambda}_s)^{-1}} + 1\right) t 2^{-i}/3}\right) \\
&\qquad \text{(By Bernstein's inequality for martingale differences (Freedman, 1975))} \\
\leq& 2\exp\left(-\frac{2^{-2i} t^2/8}{\text{term I}}\right),
\end{aligned} \tag{D.4}
$$

$$
\begin{aligned}
\text{where term I} =& 2\sum_{i'=1}^{i} \sum_{s=n_{i'-1}+1}^{n_{i'}} \|x_1 - x_k\|^2_{A(\bar{\lambda}_s)^{-1}} + 2\sum_{s=n_i+1}^{t} \|x_1 - x_k\|^2_{A(\bar{\lambda}_s)^{-1}} \\
&+ \left(\sqrt{d}\max_{s\in[1:t]} \|x_1 - x_k\|_{A(\bar{\lambda}_s)^{-1}} + 1\right) \cdot \frac{t 2^{-i}}{3}.
\end{aligned}
$$

Here, to use Bernstein's inequality for martingale differences in the inequality (a) above, we need to bound the variance and magnitude of $\left(\theta - A(\lambda_s)^{-1} x_s r_s\right)^\top (x_1 - x_k)$ condition on λ_s.[3] In particular, we have

$$
\begin{aligned}
\left|\left(\theta - A(\lambda_s)^{-1} x_s r_s\right)^\top (x_1 - x_k)\right| &\leq \left|(x_1 - x_k)^\top A(\lambda_s)^{-1} x_s\right| + \Delta_k \\
&\leq \|x_1 - x_k\|_{A(\lambda_s)^{-1}} \|x_s\|_{A(\lambda_s)^{-1}} + 2 \\
&\leq 2\sqrt{d} \|x_1 - x_k\|_{A(\bar{\lambda}_s)^{-1}} + 2. \\
&\qquad \text{(Since } \lambda_s = (\bar{\lambda}_s + \lambda^*)/2 \text{ and } \lambda \mapsto \|x_1 - x_k\|^2_{A(\lambda)^{-1}} \text{ is convex in } \lambda)
\end{aligned}
$$

$$
\mathbb{E}\left[\left(\left(\theta - A(\lambda_s)^{-1} x_s r_s\right)^\top (x_1 - x_k)\right)^2 \mid \lambda_s\right]
$$

[2] Otherwise, ξ_i is vacuously true and $\mathbb{P}(\xi_i^c) = 0$.

[3] Since IPS estimator is unbiased and λ_s is determined by the history prior to time s, we have $\mathbb{E}\left[\left(\theta - A(\lambda_s)^{-1} x_s r_s\right)^\top (x_1 - x_k) \mid \mathcal{H}_{s-1}\right] = 0$, which implies that it is a martingale difference sequence.

$$
\begin{aligned}
&\leq \mathbb{E}\left[\left((x_1 - x_k)^\top A(\lambda_s)^{-1} x_s\right)^2 \mid \lambda_s\right] \\
&= (x_1 - x_k)^\top A(\lambda_s)^{-1} \mathbb{E}\left[x_s x_s^\top \mid \lambda_s\right] A(\lambda_s)^{-1}(x_1 - x_k) \\
&= \|x_1 - x_k\|^2_{A(\lambda_s)^{-1}} \\
&\leq 2\|x_1 - x_k\|^2_{A(\bar{\lambda}_s)^{-1}}. && \text{(Since } \lambda_s = (\bar{\lambda}_s + \lambda^*)/2\text{)}
\end{aligned}
$$

Single-term error probability. Now, we need to use the property of the subroutine RAGE-Elimination (Line 5.3 of Algorithm 5.2) that generates λ_s. That is, by Lemma 56, since $x_k \in \mathcal{A}_i \subseteq \mathcal{A}_{i'}$ for $i' \leq i$ and $m \geq i_0$, for $s \in [n_{i'-1}+1, n_{i'}]$, we have $\|x_1 - x_k\|^2_{A(\bar{\lambda}_s)^{-1}} \leq m \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x,x' \in \mathcal{A}_{i'-2}} \|x - x'\|^2_{A(\lambda)^{-1}} \stackrel{\text{def}}{=} m f(\mathcal{A}_{i'-2})$. Thus, we have

$$
\begin{aligned}
&\mathbb{P}(|\Delta_k - \widehat{\Delta}_k^{(t)}| \geq 2^{-i}/2) \\
\leq& 2\exp\left(-\frac{2^{-2i}t^2/8}{2m\sum_{i'=1}^{i}(n_{i'} - n_{i'-1})f(\mathcal{A}_{i'-2}) + 2m(t - n_i)f(\mathcal{A}_{i-1}) + (m\sqrt{df(\mathcal{X})} + 1)t2^{-i}/3}\right) \\
\leq& 2\exp\left(-\frac{2^{-2i}n_i^2/8}{2m\sum_{i'=1}^{i}(n_{i'} - n_{i'-1})f(\mathcal{A}_{i'-2}) + (m\sqrt{df(\mathcal{X})} + 1)n_i 2^{-i}/3}\right),
\end{aligned}
$$

where the last inequality above holds because of $t \geq n_i$ and a simple fact that $t \mapsto \frac{t^2}{at+b}$ is an increasing function when $t \geq 0$ if $a > 0$ and $b > 0$.

Final error probability. Then, with the union bound over all $t \geq n_i$ and $k \in [K]$, it holds for any $0 < n_1 \leq n_2 \ldots \leq n_i \leq T$ that

$$
\begin{aligned}
\mathbb{P}\left(\xi_i^c \,\middle|\, \bigcap_{j=1}^{i-1} \xi_j\right) &\leq 2KT\exp\left(-\frac{2^{-2i}n_i^2/8}{2m\sum_{i'=1}^{i}(n_{i'} - n_{i'-1})f(\mathcal{A}_{i'-2}) + (m\sqrt{df(\mathcal{X})} + 1)n_i 2^{-i}/3}\right) \\
&\leq 2KT\max_{k\in[K]}\exp\left(-\frac{3n_{\bar{i}(k)}^2\Delta_k^2}{48m\sum_{i'=1}^{\bar{i}(k)}(n_{i'} - n_{i'-1})f(\mathcal{A}_{i'-2}) + 8(m\sqrt{df(\mathcal{X})} + 1)n_{\bar{i}(k)}\Delta_k}\right),
\end{aligned}
$$

where $\bar{i}(k) = \max\left\{i \in [i_0 - 1] \mid \Delta_k \leq 2^{-i}\right\}$. Here, the last inequality use the same simple fact that $t \mapsto \frac{t^2}{at+b}$ is an increasing function when $t \geq 0$ if $a > 0$ and $b > 0$.

With values of $0 = n_0 < n_1 \leq n_2 \leq \cdots \leq n_{i_0} = T$, we can define $a_i = \frac{n_i}{T}$, which implies $0 = a_0 < a_1 \leq a_2 \leq \cdots \leq a_{i_0} = 1$. Since the choice of values $\mathbf{a} \in \mathcal{D}$ is arbitrary, the final error probability can be bounded as

$$
\begin{aligned}
\mathbb{P}\left(J_T \neq 1\right) &\leq \sum_{i=1}^{i_0}\mathbb{P}\left(\xi_j^c \,\middle|\, \bigcap_{j=1}^{i-1}\xi_j\right) \\
&\leq 2i_0KT\min_{\mathbf{a}\in\mathcal{D}}\max_{k\in[K]}\exp\left(-\frac{3Ta_{i(k)}^2\Delta_k^2}{48m\sum_{i'=1}^{\bar{i}(k)}(a_{i'} - a_{i'-1})f(\mathcal{A}_{i'-2}) + 8(m\sqrt{df(\mathcal{X})} + 1)a_{i(k)}\Delta_k}\right),
\end{aligned}
$$

which completes the proof $\square$

Properties of RAGE-Elimination

In this section, we prove some properties of the RAGE-Elimination algorithm that will be useful for proving Theorem 26.

Lemma 54. *Assume $t \geq n_i$. Then, under $\bigcap_{j=1}^{i-1} \xi_j$, when running RAGE-Elimination (line 5.3 in Algorithm 5.2), it holds that*

$$\mathcal{X}_t^{(i+1)} \subseteq \left\{x \in \mathcal{X} \mid \widehat{\Delta}_x^{(t)} \leq 2^{-i}\right\} \subseteq \mathcal{A}_i.$$

Proof. To show $\mathcal{X}_t^{(i+1)} \subseteq \left\{x \in \mathcal{X} \mid \widehat{\Delta}_x^{(t)} \leq 2^{-i}\right\}$, let $x_{\widehat{(1)}_t} = \arg\max_{x \in \mathcal{X}} \left\langle \widehat{\theta}_t, x \right\rangle$. Then, for some arm x, if we have $\left\langle \widehat{\theta}^{(t)}, x_{\widehat{(1)}_t} - x \right\rangle \leq 2^{-i}$, it holds that

$$\left\langle \widehat{\theta}^{(t)}, x_1 - x \right\rangle = \underbrace{\left\langle \widehat{\theta}^{(t)}, x_1 - x_{\widehat{(1)}_t} \right\rangle}_{\leq 0} + \underbrace{\left\langle \widehat{\theta}^{(t)}, x_{\widehat{(1)}_t} - x \right\rangle}_{\leq 2^{-i}} \leq 2^{-i},$$

which implies $x \in \{x \in \mathcal{X} \mid \widehat{\Delta}_x^{(t)} \leq 2^{-i}\}$.

To show $\left\{x \in \mathcal{X} \mid \widehat{\Delta}_x^{(t)} \leq 2^{-i}\right\} \subseteq \mathcal{A}_i$, let $\widehat{\Delta}_x^{(t)} \leq 2^{-i}$ for some x and assume for the sake of a contradiction that $\Delta_x > 2 \cdot 2^{-i}$. As $\Delta_x > 2 \cdot 2^{-i}$, there must exist $\tilde{i} \leq i-1$ such that $2^{-\tilde{i}} < \Delta_x \leq 2 \cdot 2^{-\tilde{i}}$. Then $|\Delta_x - \widehat{\Delta}_x^{(t)}| < 2^{-\tilde{i}}/2$ since event $\xi_{\tilde{i}}$ holds. Meanwhile, we have $\widehat{\Delta}_x^{(t)} \leq 2^{-i} \leq 2^{-\tilde{i}}/2$ since $\tilde{i} \leq i-1$. Now, this leads to the contradiction

$$2^{-\tilde{i}}/2 = 2^{-\tilde{i}} - 2^{-\tilde{i}}/2 \leq \Delta_x - \widehat{\Delta}_x^{(t)} \leq |\Delta_x - \widehat{\Delta}_j^{(t)}| < 2^{-\tilde{i}}/2.$$

Thus, under $\bigcap_{j=1}^{i-1} \xi_j$, we have

$$\left\{x \in \mathcal{X} \mid \widehat{\Delta}_x^{(t)} \leq 2^{-i}\right\} \subseteq \left\{x \in \mathcal{X} \mid \Delta_x \leq 2 \cdot 2^{-i}\right\} = \mathcal{A}_i.$$

□

Lemma 55. *Assume $t \geq n_i$. Then, under $\bigcap_{j=1}^{i-1} \xi_j$, when running RAGE-Elimination, if $x \in \mathcal{A}_i$, then $x \in \mathcal{X}_t^{(i-1)}$.*

Proof. If $x \in \mathcal{A}_i$, then $\langle \theta, x_1 - x \rangle \leq 2 \cdot 2^{-i}$. Again, let $x_{\widehat{(1)}_t} = \arg\max_{x \in \mathcal{X}} \left\langle \widehat{\theta}_t, x \right\rangle$ and we have

$$\begin{aligned}
\left\langle \widehat{\theta}_t, \hat{x}_1^{(t)} - x \right\rangle &= \left\langle \widehat{\theta}_t, x_{\widehat{(1)}_t} - x_1 \right\rangle + \left\langle \widehat{\theta}_t, x_1 - x \right\rangle \\
&= \left\langle \widehat{\theta}_t, x_{\widehat{(1)}_t} - x_1 \right\rangle + \left\langle \widehat{\theta}_t - \theta, x_1 - x \right\rangle + \underbrace{\langle \theta, x_1 - x \rangle}_{\leq 2 \cdot 2^{-i}} \\
&\leq \left\langle \widehat{\theta}_t, x_{\widehat{(1)}_t} - x_1 \right\rangle + |\widehat{\Delta}_x^{(t)} - \Delta_x| + 2 \cdot 2^{-i} \\
&\leq \left\langle \widehat{\theta}_t, x_{\widehat{(1)}_t} - x_1 \right\rangle + 2^{-i} + 2 \cdot 2^{-i} && \text{(Since } \xi_{i-1} \text{ holds)} \\
&= -\widehat{\Delta}_{x_{\widehat{(1)}_t}}^{(t)} + 2^{-i} + 2 \cdot 2^{-i} \\
&\leq 2^{-i} + 2^{-i} + 2 \cdot 2^{-i}
\end{aligned}$$

$$=4 \cdot 2^{-i}.$$

The last inequality above holds because under $\bigcap_{j=1}^{i-1} \xi_j$, by Lemma 54, we have $x_{\widehat{(1)}_t} \in \mathcal{A}_i$, meaning that $|\widehat{\Delta}^{(t)}_{x_{\widehat{(1)}_t}} - \Delta_{x_{\widehat{(1)}_t}}| < 2^{-i} \implies \widehat{\Delta}^{(t)}_{x_{\widehat{(1)}_t}} > \Delta_{x_{\widehat{(1)}_t}} - 2^{-i} > -2^{-i}$. □

Lemma 56. *Assume $t \geq n_i$ and $\bigcap_{j=1}^{i-1} \xi_j$ holds. When running RAGE-Elimination, If $x_k \in \mathcal{A}_i$, then*

$$\|x_1 - x_k\|^2_{A(\bar{\lambda}_t)^{-1}} \leq m \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x,x' \in \mathcal{A}_{i-2}} \|x - x'\|^2_{A(\lambda)^{-1}}.$$

Proof. By Lemma 55, we have $x_1, x_k \in \mathcal{A}_i \implies x_1, x_k \in \mathcal{X}_t^{(i-1)}$, which means that $\left|\mathcal{X}_t^{(i-1)}\right| \geq 2$ and $\bar{\lambda}_t = \frac{1}{i_t} \sum_{i'=1}^{i_t} \lambda_t^{(i')}$ for some i_t satisfying $i - 1 \leq i_t \leq m$. Thus, We have

$$\begin{aligned}
\|x_1 - x_k\|^2_{A(\bar{\lambda}_t)^{-1}} \leq & m \|x_1 - x_k\|^2_{A\left(\lambda_t^{(i-1)}\right)^{-1}} \\
\leq & m \max_{x,x' \in \mathcal{X}_t^{(i-1)}} \|x - x'\|^2_{A\left(\lambda_t^{(i-1)}\right)^{-1}} && \text{(Since } x_1, x_k \in \mathcal{X}_{i-1}^{(t)}\text{)} \\
\overset{\text{(i)}}{\leq} & m \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x,x' \in \mathcal{A}_{i-2}} \|x - x'\|^2_{A(\lambda)^{-1}}.
\end{aligned}$$

Here, the above inequality (i) holds because by Lemma 54, we have $\mathcal{X}_t^{(i-1)} \subseteq \mathcal{A}_{i-2}$ and by algorithm construction, we have $\lambda_t^{(i-1)} \in \operatorname{argmin}_{\lambda \in \triangle_{\mathcal{X}}} \max_{x,x' \in \mathcal{X}_t^{(i-1)}} \|x - x'\|^2_{A(\lambda)^{-1}}$. □

Simplified Stationary Complexity and its Relation to Multi-armed Bandits

In this section, we simplify the complexity of Algorithm 5.2 obtained in Theorem 26 by appropriately choosing values $\mathbf{a} \in \mathcal{D}$. In particular, we have the following theorem.

Theorem 27. *For $\overline{H}_{\text{P1-RAGE}}(\theta)$ defined in equation* (D.3)*, we have*

$$\overline{H}_{\text{P1-RAGE}}(\theta) \leq \frac{1024 m i_0}{\Delta_1} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_1} \frac{\|x - x_1\|^2_{A(\lambda)^{-1}}}{\Delta_x} + \frac{16 m \sqrt{d}}{3\Delta_1} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_1} \|x - x_1\|_{A(\lambda)^{-1}} + \frac{1}{3\Delta_1}.$$

Proof. For $i \in \{1, \ldots, i_0 - 1\}$, we take $a_i = \frac{\Delta_1}{\Delta_{\bar{k}(i)}}$, where $\bar{k}(i) = \min\left\{k \in [K] \mid \Delta_k \geq \frac{2^{-i}}{2}\right\}$. Then, since $\bar{i}(k) = \max\left\{i \in [i_0 - 1] \mid \Delta_k \leq 2^{-i}\right\}$, for any $k \in [K]$, we have $\frac{2^{-\bar{i}(k)}}{2} \leq \Delta_{\bar{k}(\bar{i}(k))} \leq \Delta_k$, which further implies

$$a_{\bar{i}(k)} \Delta_k = \frac{\Delta_1}{\Delta_{\bar{k}(\bar{i}(k))}} \cdot \Delta_k \geq \Delta_1.$$

Then, for $\overline{H}_{\text{P1-RAGE}}(\theta)$ (defined in equation (D.3)), we have

$$\begin{aligned}
\overline{H}_{\text{P1-RAGE}}(\theta) &\leq \max_{k \in [K]} \left\{ \frac{16 m \sum_{i'=1}^{\bar{i}(k)} (a_{i'} - a_{i'-1}) f(\mathcal{A}_{i'-2})}{a^2_{\bar{i}(k)} \Delta_k^2} + \frac{8(m\sqrt{df(\mathcal{X})} + 1)}{3 a_{\bar{i}(k)} \Delta_k} \right\} \\
&\leq \frac{16m}{\Delta_1} \max_{k \in [K]} \left\{ \frac{f(\mathcal{A}_{-1})}{\Delta_{\bar{k}(1)}} + \sum_{i'=2}^{\bar{i}(k)} \left(\frac{1}{\Delta_{\bar{k}(i')}} - \frac{1}{\Delta_{\bar{k}(i'-1)}} \right) f(\mathcal{A}_{i'-2}) \right\} + \frac{8(m\sqrt{df(\mathcal{X})} + 1)}{3\Delta_1}. \\
& && \text{(Since } a_0 = 0 \text{ by definition)}
\end{aligned}$$

For the second term, using the definition of $f(\mathcal{X})$, we simply have

$$\begin{aligned}\frac{8(m\sqrt{df(\mathcal{X})}+1)}{3\Delta_1} =& \frac{8m\sqrt{d}}{3\Delta_1}\inf_{\lambda\in\triangle_\mathcal{X}}\max_{x,x'\in\mathcal{X}}\left\|x-x_1+x_1-x'\right\|_{A(\lambda)^{-1}}+\frac{1}{3\Delta_1}\\ \leq& \frac{16m\sqrt{d}}{3\Delta_1}\inf_{\lambda\in\triangle_\mathcal{X}}\max_{x\neq x_1}\|x-x_1\|_{A(\lambda)^{-1}}+\frac{1}{3\Delta_1}. \end{aligned} \tag{D.5}$$

For the first term, by fixing arm index $k\in[K]$ and defining $j\in\arg\max_{\ell\in[\bar{i}(k)]}\frac{f(\mathcal{A}_{\ell-2})}{\Delta_{\bar{k}(\ell)}}$, we have

$$\begin{aligned}&\frac{f(\mathcal{A}_{-1})}{\Delta_{\bar{k}(1)}}+\sum_{i'=2}^{\bar{i}(k)}\left(\frac{1}{\Delta_{\bar{k}(i')}}-\frac{1}{\Delta_{\bar{k}(i'-1)}}\right)f(\mathcal{A}_{i'-2})\\ =&\frac{f(\mathcal{A}_{\bar{i}(k)-2})}{\Delta_{\bar{k}(\bar{i}(k))}}+\sum_{i'=1}^{\bar{i}(k)-1}\frac{f(\mathcal{A}_{i'-2})-f(\mathcal{A}_{i'-1})}{\Delta_{\bar{k}(i')}}\\ \overset{\text{(a)}}{\leq}&\frac{f(\mathcal{A}_{j-2})}{\Delta_{\bar{k}(j)}}\left(1+\sum_{i'=1}^{\bar{i}(k)-1}\frac{f(\mathcal{A}_{i'-2})-f(\mathcal{A}_{i'-1})}{f(\mathcal{A}_{i'-2})}\right)\\ \leq&\bar{i}(k)\frac{f(\mathcal{A}_{j-2})}{\Delta_{\bar{k}(j)}} && \text{(Since } f(\mathcal{A}_{i'-2})\geq f(\mathcal{A}_{i'-1})\text{)}\\ \leq& i_0\max_{\ell\in[\bar{i}(k)]}\frac{f(\mathcal{A}_{\ell-2})}{\Delta_{\bar{k}(\ell)}} && \text{(Since } \bar{i}(k)\leq i_0 \text{ for any } k\in[K]\text{)}\\ =& i_0\max_{\ell\in[\bar{i}(k)]}\inf_{\lambda\in\triangle_\mathcal{X}}\max_{x,x'\in\mathcal{A}_{\ell-2}}\frac{\|x-x'\|^2_{A(\lambda)^{-1}}}{\Delta_{\bar{k}(\ell)}}\\ \leq& i_0\inf_{\lambda\in\triangle_\mathcal{X}}\max_{\ell\in[\bar{i}(k)]}\max_{x,x'\in\mathcal{A}_{\ell-2}}\frac{\|x-x'\|^2_{A(\lambda)^{-1}}}{\Delta_{\bar{k}(\ell)}} && \text{(By the weak duality inequality)}\\ \leq& 64i_0\inf_{\lambda\in\triangle_\mathcal{X}}\max_{\ell\in[\bar{i}(k)]}\max_{x\in\mathcal{A}_{\ell-2},x\neq x_1}\frac{\|x-x_1\|^2_{A(\lambda)^{-1}}}{16\Delta_{\bar{k}(\ell)}} && \text{(By reasoning similar to equation (D.5))}\\ \overset{\text{(b)}}{\leq}&64i_0\inf_{\lambda\in\triangle_\mathcal{X}}\max_{\ell\in[\bar{i}(k)]}\max_{x\in\mathcal{A}_{\ell-2},x\neq x_1}\frac{\|x-x_1\|^2_{A(\lambda)^{-1}}}{\Delta_x}\\ \leq&64i_0\inf_{\lambda\in\triangle_\mathcal{X}}\max_{x\neq x_1}\frac{\|x-x_1\|^2_{A(\lambda)^{-1}}}{\Delta_x}.\end{aligned}$$

Here, the inequality (a) above holds because $f(\mathcal{A}_{i'-2})\geq f(\mathcal{A}_{i'-1})$ and by definition of j, we have $\frac{f(\mathcal{A}_{\ell-2})}{\Delta_{\bar{k}(\ell)}}\leq\frac{f(\mathcal{A}_{j-2})}{\Delta_{\bar{k}(j)}}$. The inequality (b) above holds because by definitions of $\bar{k}(\ell)=\min\left\{k\in[K]\mid\Delta_k\geq\frac{2^{-i}}{2}\right\}$ and $\mathcal{A}_{\ell-2}=\left\{x\in\mathcal{X}\mid\Delta_x\leq 2\cdot 2^{-(\ell-2)}\right\}$, we have $16\Delta_{\bar{k}(\ell)}\geq\Delta_x$ for any $x\in\mathcal{A}_{\ell-2}$.

Therefore, by plugging the bound of both terms back, we have

$$\overline{H}_{\text{P1-RAGE}}(\theta)\leq\frac{1024mi_0}{\Delta_1}\inf_{\lambda\in\triangle_\mathcal{X}}\max_{x\neq x_1}\frac{\|x-x_1\|^2_{A(\lambda)^{-1}}}{\Delta_x}+\frac{16m\sqrt{d}}{3\Delta_1}\inf_{\lambda\in\triangle_\mathcal{X}}\max_{x\neq x_1}\|x-x_1\|_{A(\lambda)^{-1}}+\frac{1}{3\Delta_1}.$$

□

In the following corollary, we show that the above simplified complexity is in a same order (up to logarithmic factors) of H_{BOB} proposed in Abbasi-Yadkori et al. (2018).

Corollary 5. *In multi-armed bandits, meaning $d = K$ and $\mathcal{X} = \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$, for $H_{\text{P1-RAGE}}(\theta)$ (defined in equation* (5.4)*), if $m = i_0$, we then have*

$$H_{\text{P1-RAGE}}(\theta) \leq \frac{2i_0\left(i_0 \log(2K) + 1\right)}{\Delta_{(1)}} \max_{k \in [K]} \frac{k}{\Delta_{(k)}} = 2i_0\left(i_0 \log(2K) + 1\right) H_{\mathrm{BOB}}(\theta).$$

Proof. When in multi-armed bandits, for the first term in $H_{\mathsf{P1\text{-}RAGE}}(\theta)$, we have

$$\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \frac{\left\|x - x_{(1)}\right\|^2_{A(\lambda)^{-1}}}{\Delta_x} \leq 2 \sum_{k=1}^{K} \frac{1}{\Delta_k} \leq 2 \log(2K) \max_{k \in [K]} \frac{k}{\Delta_{(k)}},$$

where the first inequality above comes from Soare et al. (2014) and the second inequality comes from Audibert et al. (2010). For the second term in $H_{\mathsf{P1\text{-}RAGE}}(\theta)$, we have

$$\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \left\|x - x_{(1)}\right\|_{A(\lambda)^{-1}} = \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{k \neq (1)} \sqrt{\frac{1}{\lambda_{(1)}} + \frac{1}{\lambda_k}} = \sqrt{2K},$$

which then gives us $\frac{\sqrt{K} \cdot \sqrt{2K}}{\Delta_{(1)}} \leq \frac{2K}{\Delta_{(1)} \Delta_{(K)}} \leq \frac{2}{\Delta_{(1)}} \max_{k \in [K]} \frac{k}{\Delta_{(k)}}$.

Finally, by plugging these inequalities back into $H_{\mathsf{P1\text{-}RAGE}}(\theta)$ (defined in equation (5.4)), we have

$$\begin{aligned}
H_{\mathsf{P1\text{-}RAGE}}(\theta) =& \frac{mi_0}{\Delta_{(1)}} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \frac{\left\|x - x_{(1)}\right\|^2_{A(\lambda)^{-1}}}{\Delta_x} + \frac{m\sqrt{d}}{\Delta_{(1)}} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \neq x_{(1)}} \left\|x - x_{(1)}\right\|_{A(\lambda)^{-1}} \\
\leq& \frac{2i_0^2 \log(2K)}{\Delta_{(1)}} \max_{k \in [K]} \frac{k}{\Delta_{(k)}} + \frac{2i_0}{\Delta_{(1)}} \max_{k \in [K]} \frac{k}{\Delta_{(k)}} \\
=& \frac{2i_0\left(i_0 \log(2K) + 1\right)}{\Delta_{(1)}} \max_{k \in [K]} \frac{k}{\Delta_{(k)}}.
\end{aligned}$$

□

Approximate BAI of Algorithm 5.2

Corollary 6. *Fix arm set $\mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = K$ and budget T. For a stationary environment with unknown parameter θ, if $m \geq i_0(\epsilon) = \lceil \log_2(1/\epsilon) \rceil + 1$ for some $\epsilon \geq \Delta_1$, then there exists absolute constant $c > 0$ such that the error probability of P1-RAGE satisfies*

$$\mathbb{P}_\theta\left(J_T \notin \mathcal{A}(\epsilon)\right) \leq 2i_0(\epsilon) KT \exp\left(-\frac{cT}{H_{\text{P1-RAGE}}(\theta, \epsilon)}\right),$$

where $\mathcal{A}(\epsilon) = \{x \in \mathcal{X} \mid \Delta_x \leq \epsilon\}$ and $H_{\text{P1-RAGE}}(\theta, \epsilon)$ is defined as replacing i_0 by $i_0(\epsilon)$ in $H_{\text{P1-RAGE}}(\theta)$ (defined in Eq. (D.3)*).*

Proof. The proof is the same as Theorem 15 through simply replacing i_0 by $i_0(\epsilon)$. □

D.3.2 Non-stationary Environments

In this section, we prove the error probability of Algorithm 5.2 in general non-stationary environments.

Theorem 28. *Fix time horizon T, arm set $\mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = K$ and arbitrary unknown parameters $\{\theta_t\}_{t=1}^T$. If we run Algorithm 5.2 in this non-stationary environment and obtain x_{J_T}, then it holds that*

$$\mathbb{P}_{\overline{\theta}_T}\left(J_T \neq (1)\right) \leq K \exp\left(-\frac{3T\Delta_{(1)}^2}{64d}\right).$$

Proof. The proof will basically resemble the one for Theorem 14. In particular, by the same reasoning to obtain equation D.2, we have

$$\mathbb{P}\left(J_T \neq (1)\right) \leq \mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) + \sum_{k=2}^{K} \mathbb{P}\left(x_{(k)}^\top \widehat{\theta}_T - x_{(k)}^\top \overline{\theta}_T \geq \frac{\Delta_{(k)}}{2}\right),$$

$$\text{where } \mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) = \mathbb{P}\left(\sum_{t=1}^{T} x_{(1)}^\top \left(A(\lambda_t)^{-1} x_t r_t - \theta_t\right) \leq -\frac{T\Delta_{(1)}}{2}\right).$$

Since $\lambda_t = \frac{\bar{\lambda}_t + \lambda^*}{2}$ and $\lambda \mapsto \|x\|^2_{A(\lambda)^{-1}}$ is convex in λ, to use the Berstein's inequality for martingale differences (Freedman, 1975), we have

$$\left|x_{(1)}^\top \left(A(\lambda_t)^{-1} x_t r_t - \theta_t\right)\right| \leq 2\left\|x_{(1)}\right\|_{A(\lambda^*)^{-1}} \|x_t\|_{A(\lambda^*)^{-1}} + 2 \leq 2d + 2 \leq 4d,$$

$$\mathbb{E}\left[\left(x_{(1)}^\top \left(A(\lambda_t)^{-1} x_t r_t - \theta_t\right)\right)^2 \mid \lambda_t\right] = \left\|x_{(1)}\right\|^2_{A(\lambda_t)^{-1}} \leq 2\left\|x_{(1)}\right\|^2_{A(\lambda^*)^{-1}} \leq 2d.$$

Therefore, we have

$$\mathbb{P}\left(x_{(1)}^\top \widehat{\theta}_T - x_{(1)}^\top \overline{\theta}_T \leq -\frac{\Delta_{(1)}}{2}\right) \leq \exp\left(-\frac{T\Delta_{(1)}^2/8}{2d + 2d\Delta_{(1)}/3}\right) \leq \exp\left(-\frac{3T\Delta_{(1)}^2}{64d}\right).$$

By applying the same inequality to other terms, we have

$$\mathbb{P}\left(J_T \neq (1)\right) \leq K \exp\left(-\frac{3T\Delta_{(1)}^2}{64d}\right).$$

□

D.4 Implementation Details and Additional Experiments

In this section, we provide more implementation details and additional experiment results. Experiments are executed through Python 3.10 and paralleled by a Mac M1 Pro chip with 6 cores.

First, we notice that an algorithm for stationary environments usually determines a batch of arms to pull at once during each epoch, while in non-stationary environment, the order of pulling these arms will affect the rewards it will receive. Therefore, when applying stationary algorithms (Peace and OD-LinBAI) into a non-stationary environment, we use a random permutation to determine the order of pulling for each batch of arms.

When implementing P1-RAGE, to be computationally efficient, we update λ_t in the same frequency as P1-Peace, which is summarized in Algorithm D.1. We take $m = 15$ for P1-RAGE, which, based on Theorem 15, is valid as long as $\Delta_{(1)} \geq 2^{-13} \approx 1.22 \times 10^{-4}$. Furthermore, when implementing Peace, for simplicity, we use $\inf_{\lambda \in \triangle_{\mathcal{X}}} \rho(\mathcal{Z}, \lambda)$, defined in equation (D.1), to replace all $\gamma(\mathcal{Z})$ used in Katz-Samuels et al. (2020). Since the paper of OD-LinBAI does not provide code, we implement it based on the pseudocode in Yang and Tan (2022). Finally, we use Frank-Wolfe algorithm with stepsize $\frac{1}{2(i+2)}$ in i-th iteration to solve all optimization problems in a form of $\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}} \|y\|^2_{A(\lambda)^{-1}}$.

As for code snippets reference, we use part of the code from Katz-Samuels et al. (2020) to implement the rounding procedure used in Peace[4] and part of the code from Fiez et al. (2019) to generate the base stationary instance for the multivariate testing example.[5] We also use code from Xu et al. (2018) to preprocess the Yahoo! Webscope dataset.[6]

D.4.1 Additional Experiments

Here, we provide experiment results on some additional examples to corroborate our theoretical findings.

Malicious non-stationary example Because of the nature of arm elimination, algorithms designed for stationary environment can fail easily in some malicious non-stationary environments. Here, we pick the same $\mathcal{X}$ as Soare et al. (2014)'s stationary benchmark example and set $\omega = 0.5$. Then, we take

$$\theta_t = \begin{cases} \begin{bmatrix} 0 & 1 & 1 & \dots & 1 \end{bmatrix}^\top & \text{for } t = 1, \dots, \frac{T}{3}, \\ \begin{bmatrix} 2 & 0 & 0 & \dots & 0 \end{bmatrix}^\top & \text{for } t = \frac{T}{3} + 1, \dots, T. \end{cases}$$

We can see that the overall best arm is still $x_{(1)} = \mathbf{e}_1$. However, because of the θ_t in the first $1/3$ rounds, algorithms like Peace and OD-LinBAI will eliminate $\mathbf{e}_1$ in its initial phase; on the other hand, our algorithms will be robust to this non-stationarity. Here, we take $T = 10^4$ and the results are shown in right plot of Figure D.1.

Stationary multivariate testing example We also test the performance of these algorithms in multivariate testing example when there is no non-stationarity, i.e. $\theta_t = \theta^*$ for all t. Here, we also take $T = 10^4$ and the results are shown in Figure D.2. We can see that our robust algorithm P1-RAGE again performs better than G-BAI and comparably with Peace.

Non-stationary benchmark example In this example, we add non-stationarity to Soare et al. (2014)'s stationary benchmark example in a more structured instead of malicious way. In particular, we keep the arm set $\mathcal{X}$ the same, take $\omega = 0.5$ and set

$$\theta_t = \begin{bmatrix} 0.3 & 0 & 0 & \dots & -s \sin\left(\frac{2\pi t}{L}\right) + 0.5 \end{bmatrix}^\top,$$

where s is the oscillation scale and L is the oscillation period, In the first series of instances, we fix $L = 200$ and take values $m \in \{0, 1, \dots, 9\}$; in the second series of instances, we fix $m = 1$ and take values $L \in \{300, 600, \dots, 3000\}$. All non-stationary instances have the same optimal arm as their stationary

[4] No license information.

[5] Under MIT License.

[6] No license information.

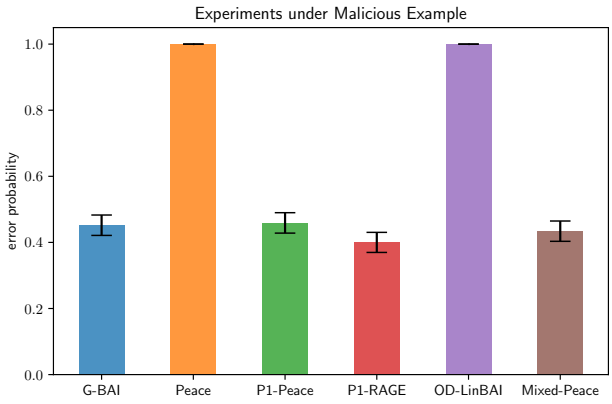


Figure D.1: The error probabilities are estimated through 1000 repeated trials and the error bars represent 95% confidence intervals.

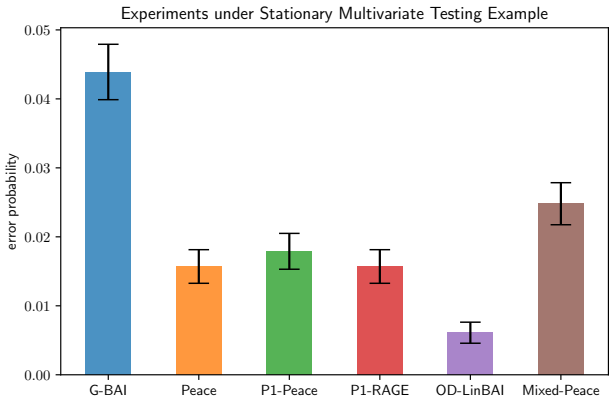


Figure D.2: The error probabilities are estimated through 10^4 repeated trials and the error bars represent 95% confidence intervals.

counterparts and we take $T = 10^4$ for all of these instances. The results are shown in Figure D.3, from which we can see similar phenomenon as in Figure 5.3. In particular, algorithms designed for stationary environments, Peace and OD-LinBAI, are very unstable in face of non-stationarity. Meanwhile, among the other four relatively robust algorithms, our algorithms P1-RAGE and P1-Peace consistently outperform the other two.

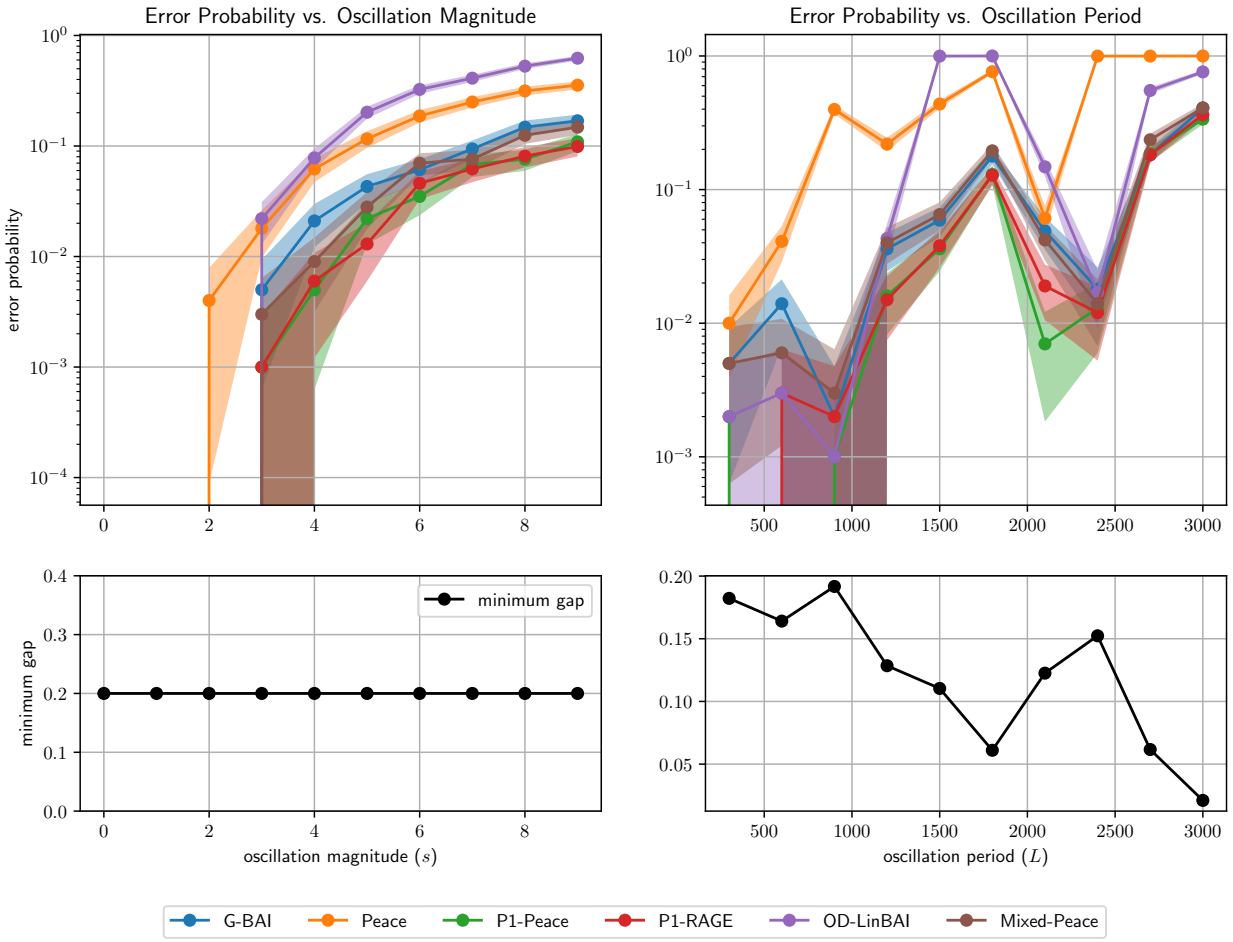


Figure D.3: The vertical axis (error probability) is in log scale. The shaded area represents the 95% confidence interval. Each error probability is estimated through 1000 repeated trials. The bottom two plots give the minimum gap $\Delta_{(1)}$ of each instance that algorithms run over

Appendix E

Appendix for Chap. 6

E.1 Lower Bounds

E.1.1 Oracle Lower Bound

Theorem 29 (Oracle Lower Bound)**.** *Let τ denote the stopping time for any $(0,\delta)$-PAC algorithm for pure exploration in safe linear bandits. Then*

$$\frac{\mathbb{E}_{\theta_*,\mu_*}[\tau]}{\log\frac{1}{2.4\delta}} \geq \min_{\lambda\in\triangle_{\mathcal{X}}} \max \left\{ \max_{z\in\mathcal{Z}\setminus z_*} \min \left\{ \frac{\|z\|^2_{A(\lambda)^{-1}}}{\mathfrak{p}(-\Delta_{\text{safe}}(z))^2}, \frac{\|z-z_*\|^2_{A(\lambda)^{-1}}}{\mathfrak{p}(\Delta(z))^2} \right\}, \frac{\|z_*\|^2_{A(\lambda)^{-1}}}{(z_*^\top\mu_*-\alpha)^2} \right\}.$$

Comparing Complexity with Theorem 16. In the single-constraint setting, the complexity of BESIDE reduces to

$$C\cdot\sup_{\widetilde{\epsilon}\geq\epsilon}\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{m|\mathcal{Z}|}{\delta})}{\left(|\Delta_{\text{safe}}(z)|+\mathfrak{p}(\Delta^{\widetilde{\epsilon}}(z))+\widetilde{\epsilon}\right)^2} + C\cdot\sup_{\widetilde{\epsilon}\geq\epsilon}\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z-z_*\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{|\mathcal{Z}|}{\delta})}{\left(\mathfrak{p}(-\Delta_{\text{safe}}(z))+\mathfrak{p}(\Delta^{\widetilde{\epsilon}}(z))+\widetilde{\epsilon}\right)^2}+C_0$$

Consider the case when $\Delta^{\widetilde{\epsilon}}(z)$ is "smooth" in $\widetilde{\epsilon}$, in the sense that $\Delta^{\widetilde{\epsilon}}(z)\geq\Delta(z)-\widetilde{\epsilon}$. This condition corresponds to the case, for example, where z_* has a large safety gap (in which case we simply have $\Delta^{\widetilde{\epsilon}}(z)=\Delta(z)$ for moderate values of $\widetilde{\epsilon}$), or where z_* might have a small safety gap, but where there are arms placed at even intervals so that, as we let the safety gap get smaller, we are always able to find better arms. Under this assumption, the complexity can be upper bounded as

$$C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{m|\mathcal{Z}|}{\delta})}{\left(|\Delta_{\text{safe}}(z)|+\mathfrak{p}(\Delta(z))\right)^2}+C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{|\mathcal{Z}|}{\delta})}{\left(\mathfrak{p}(-\Delta_{\text{safe}}(z))+\mathfrak{p}(\Delta(z))\right)^2}+C_0$$
$$\leq C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{m|\mathcal{Z}|}{\delta})}{\left(|\Delta_{\text{safe}}(z)|+\mathfrak{p}(\Delta(z))\right)^2}+C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{|\mathcal{Z}|}{\delta})}{\left(\mathfrak{p}(-\Delta_{\text{safe}}(z))+\mathfrak{p}(\Delta(z))\right)^2}+C_0$$

which can be upper bounded as

$$C\log(\tfrac{m|\mathcal{Z}|}{\delta})\cdot\left(\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z\|^2_{A(\lambda)^{-1}}}{\max\{\Delta_{\text{safe}}(z)^2,\mathfrak{p}(\Delta(z))^2\}}+\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}\setminus z_*}\frac{\|z-z_*\|^2_{A(\lambda)^{-1}}}{\max\{\mathfrak{p}(-\Delta_{\text{safe}}(z))^2,\mathfrak{p}(\Delta(z))^2\}}\right)+C_0.$$

While this does not match the lower bound of Theorem 29 exactly, it scales in a similar manner. As in Theorem 29, we pay only for the larger of the optimality gap, $\mathfrak{p}(\Delta(z))$, and safety gap $\mathfrak{p}(-\Delta_{\text{safe}}(z))$ (if the arm is unsafe). The primary difference between Theorem 29 and this complexity are the terms in the numerator—in Theorem 29, the numerator scales as $\|z - z_*\|^2_{A(\lambda)^{-1}}$ only if an arm is easier to eliminate by showing it is suboptimal, while in our complexity it could scale this way in either case.

The primary difficulty in hitting the lower bound exactly is that Theorem 29 is a *verification* lower bound. It assumes knowledge of the best arm, and is told whether every other arm has smaller safety gap (if the arm is unsafe) or optimality gap. It can therefore simply use this knowledge to focus all samples on verifying an arm is either unsafe, or suboptimal.

In practice, we do not have access to such information. Without knowing whether it is easier to eliminate an arm by showing it is unsafe or suboptimal, the best we can hope to do is to seek to estimate both the safety value and reward value of every arm, until we have estimated one well enough to show the arm is suboptimal or unsafe.

We conjecture that the lower bound of Theorem 29 is loose, and that Theorem 16 is nearly optimal. We believe the gap arises because, as noted, lower bound proof techniques, such as those proposed in (Kaufmann et al., 2016), which is what we rely on to prove Theorem 29, are lower bounding only the complexity of verifying the optimal solution. In problem settings such as ours where the *order* matters—where we will obtain a very different rate if we focus our attention on one arm versus another, to show it is safe or unsafe—such techniques appear insufficient to obtain a tight lower bound. Indeed, we conjecture that a "moderate-confidence" lower bound can be shown using techniques from (Simchowitz et al., 2017), and that such a lower bound may have a complexity nearly matching that of Theorem 16. We leave proving this for future work.

Proof of Theorem 29. Following the proof of Theorem 1 of (Fiez et al., 2019) and applying the Transportation Lemma of (Kaufmann et al., 2016), we have that any δ-PAC algorithm must satisfy

$$\sum_{x\in\mathcal{X}} \mathbb{E}[T_x] \geq \log \frac{1}{2.4\delta} \cdot \inf_{\lambda\in\triangle_{\mathcal{X}}} \frac{1}{\min_{(\theta,\mu)\in\mathcal{C}_{\text{alt}}} \sum_{x\in\mathcal{X}} \lambda_x \text{KL}(\nu_{(\theta_*,\mu_*),i}||\nu_{(\theta,\mu),i})}$$

there T_x denotes the number of pulls to arm x, and $\mathcal{C}_{\text{alt}}$ is the set of alternate instances defined in Lemma 57. As we assume that the noise is $\mathcal{N}(0,1)$, and since the noise is independent for the safety observations and reward observations, we have

$$\text{KL}(\nu_{(\theta_*,\mu_*),i}||\nu_{(\theta,\mu),i}) = \frac{1}{2}(x_i^\top(\theta_* - \theta))^2 + \frac{1}{2}(x_i^\top(\mu_* - \mu))^2.$$

Some algebra shows that

$$\sum_{x\in\mathcal{X}} \lambda_x \text{KL}(\nu_{(\theta_*,\mu_*),i}||\nu_{(\theta,\mu),i}) = \frac{1}{2}\|\theta_* - \theta\|^2_{A(\lambda)} + \frac{1}{2}\|\mu_* - \mu\|^2_{A(\lambda)}.$$

The result then follows by applying Lemma 57 to compute

$$\min_{(\theta,\mu)\in\mathcal{C}_{\text{alt}}} \frac{1}{2}\|\theta_* - \theta\|^2_{A(\lambda)} + \frac{1}{2}\|\mu_* - \mu\|^2_{A(\lambda)}.$$

□

Lemma 57. *Define the alternate set:*

$$\mathcal{C}_{\text{alt}} = \{(\theta, \mu) \text{ s.t. } \mu^\top z^* > \alpha\} \cup \{(\theta, \mu) \text{ s.t. } \exists z' \neq z^*, \mu^\top z' \leq \alpha, \theta^\top (z^* - z') \leq 0\},$$

Then the projection to the alternate is

$$\min_{(\theta,\mu)\in\mathcal{C}_{\text{alt}}} \|\theta - \theta_*\|^2_{A(\lambda)} + \|\mu - \mu_*\|^2_{A(\lambda)} = \min\left\{ \min_{z\neq z^*} \frac{\mathfrak{p}(z^\top \mu_* - \alpha)^2}{\|z\|^2_{A(\lambda)^{-1}}} + \frac{\mathfrak{p}((z_* - z)^\top \theta_*)^2}{\|z - z_*\|^2_{A(\lambda)^{-1}}}, \frac{(z_*^\top \mu_* - \alpha)^2}{\|z_*\|^2_{A(\lambda)^{-1}}} \right\}.$$

Proof. For each arm x the associated and we want to solve

$$\min_{(\theta,\mu)\in\mathcal{C}_{\text{alt}}} \|\theta - \theta_*\|^2_{\sum_{x\in\mathcal{X}} \lambda_x x x^\top} + \|\mu - \mu_*\|^2_{\sum_{x\in\mathcal{X}} \lambda_x x x^\top}.$$

To do so, we use that $\min_{x\in A\cup B} f(x) = \min_{S\in\{A,B\}} \min_{x\in S} f(x)$ on the quadratic objective by defining the sets

$$A := \{(\theta, \mu) \text{ s.t. } \mu^\top z^* > \alpha\}\ ,\ B = \{(\theta, \mu) \text{ s.t. } \exists z' \neq z^*, \mu^\top z' \leq \alpha, \theta^\top (z^* - z') \leq 0\},$$

such that their union is $A \cup B = \mathcal{C}_{\text{alt}}$.

Note that we know from (Mason et al., 2021) that

$$\min_{(\theta,\mu)\in A} \|\theta - \theta_*\|^2_{\sum_{x\in\mathcal{X}} \lambda_x x x^\top} + \|\mu - \mu_*\|^2_{\sum_{x\in\mathcal{X}} \lambda_x x x^\top} = \frac{(z_*^\top \mu_* - \alpha)^2}{\|z_*\|^2_{A(\lambda)^{-1}}}.$$

We now lift B to a set $\text{lift}(B)$ that is defined as

$$\text{lift}(B) = \{[\theta, \mu] \text{ s.t. } \exists z' \neq z^*, [\theta, \mu]^\top [(z^* - z'), 0; 0, z'] \leq [0, \alpha]\}.$$

Thus we can focus on $D_z = \{\kappa \in \mathbb{R}^{2n} \text{ s.t. } A_z \kappa \leq b\}$ where $A_z = [(z^* - z'), 0; 0, z'] \in R^{2\times 2n}$. Now we want to solve

$$\min_{z\in\mathcal{Z}\setminus\{z_*\}} \min_{\kappa\in\mathbb{R}^{2n} \text{ s.t. } A_z\kappa\leq b} \|\kappa - \kappa_*\|_\Gamma,$$

where $\Gamma = I_2 \bigotimes \left(\sum_{x\in\mathcal{X}} \lambda_x x x^\top\right)$.

Lemma 58. *The optimal solution of*

$$\min_{\kappa\in\mathbb{R}^{2n} \text{ s.t. } A\kappa\leq b} \frac{\|\kappa - \kappa_*\|_\Gamma}{2}$$

is $\kappa_0 = \kappa_* - \Gamma^{-1} A^\top (A\Gamma^{-1} A^\top)^{-1} \{A\kappa_* - b\}_+$ *and the optimal value is*

$$\frac{1}{2} \mathfrak{p}(A\kappa_* - b)^\top (A\Gamma^{-1} A^\top)^{-1} \mathfrak{p}(A\kappa_* - b),$$

where $\mathfrak{p}(\cdot)$ *is applied element-wise to* $A\kappa_* - b$.

This translate to

$$\min_{(\theta,\mu)\in B} \|\theta - \theta_*\|^2_{\sum_{x\in\mathcal{X}} \lambda_x x x^\top} + \|\mu - \mu_*\|^2_{\sum_{x\in\mathcal{X}} \lambda_x x x^\top} = \min_{z\neq z^*} \frac{\mathfrak{p}(z^\top \mu_* - \alpha)^2}{\|z\|^2_{A(\lambda)^{-1}}} + \frac{\mathfrak{p}((z_* - z)^\top \theta_*)^2}{\|z - z_*\|^2_{A(\lambda)^{-1}}},$$

and we get the desired result.

□

Proof of Lemma 58. Consider the Lagrangian

$$\mathcal{L}(\kappa, \mu) = \frac{1}{2}(\kappa - \kappa_*)^\top \Gamma(\kappa - \kappa_*) + \mu^\top (A\kappa - b)$$
$$\mathcal{L}(\kappa', \mu) = \frac{1}{2}\kappa'^\top \Gamma \kappa' + \mu^\top (A\kappa' - b + A\kappa_*)$$

minimized at $\kappa'_0 = -\Gamma^{-1}A^\top \mu$. We have

$$\begin{aligned}\max_{\mu \geq 0} \min_{\kappa'} \mathcal{L}(\kappa', \mu) &= \max_{\mu \geq 0} \frac{1}{2}(\Gamma^{-1}A^\top \mu)^\top \Gamma(\Gamma^{-1}A^\top \mu) + \mu^\top(-A\Gamma^{-1}A^\top \mu - b + A\kappa_*) \\ &= \max_{\mu \geq 0} -\frac{1}{2}\mu^\top A\Gamma^{-1}A^\top \mu + \mu^\top(A\kappa_* - b)\end{aligned}$$

maximized at $\mu_0 = (A\Gamma^{-1}A\top)^{-1}\{A\kappa_* - b\}_+$ where $\{[b_1, b_2]\}_+ = [\max\{b_1, 0\}, \max\{b_2, 0\}]$. Plugging μ_0 back in the solution κ'_0, we get the solution κ_0

$$\kappa_0 = \kappa_* - \Gamma^{-1}A^\top (A\Gamma^{-1}A^\top)^{-1}\{A\kappa_* - b\}_+$$

and the optimal value follows. □

E.1.2 Proof of Proposition 1

Proof of Proposition 1.

Proof for $\mathcal{I}^1$. Fix $\alpha \in (0, 0.1)$ and consider the following instance with $m = 1$:

$$\begin{aligned}&\mathcal{X} = \{e_1, e_2\}, \quad \mathcal{Z} = \{z_1, z_2\}, \quad z_1 = [1/4, 1/2], \quad z_2 = [3/4, 1/2 + \alpha] \\ &\theta_* = e_1, \quad \mu_* = [0, 1], \quad \gamma = 1/2 + \alpha/2.\end{aligned}$$

On this example, z_1 is safe and z_2 is unsafe with $\Delta_{\text{safe}}(z_2) = -\alpha/2$.

Let $A(\lambda) = \lambda_1 e_1 e_1^\top + \lambda_2 e_2 e_2^\top$ denote the design matrix. Then the allocation that minimizes $\mathcal{XY}_{\text{diff}}$:

$$\max_{z, z' \in \mathcal{Z}} \|z - z'\|^2_{A(\lambda)^{-1}} = \frac{1}{4\lambda_1} + \frac{\alpha^2}{\lambda_2}$$

is

$$\lambda_1 = \frac{1}{1 + 2\alpha}, \quad \lambda_2 = \frac{2\alpha}{1 + 2\alpha}.$$

Denote this allocation as $\widetilde{\lambda}$.

Applying the Transportation Lemma of (Kaufmann et al., 2016), this implies that any δ-PAC strategy must have

$$\mathbb{E}[T_1]\text{KL}(\nu_{(\theta_*, \mu_*), 1} || \nu_{(\theta, \mu), 1}) + \mathbb{E}[T_2]\text{KL}(\nu_{(\theta_*, \mu_*), 2} || \nu_{(\theta, \mu), 2}) \geq \log \frac{1}{2.4\delta}$$

for all $(\theta, \mu) \in \mathcal{C}_{\text{alt}}$, where $\mathcal{C}_{\text{alt}}$ is defined as in Lemma 57. If a learner plays $\widetilde{\lambda}$ for T steps, they will have $\mathbb{E}[T_1] = \lambda_1 \mathbb{E}[T], \mathbb{E}[T_2] = \lambda_2 T$. In this case, the above can be rewritten as

$$\mathbb{E}[T] \geq \log \frac{1}{2.4\delta} \cdot \frac{1}{\lambda_1 \text{KL}(\nu_{(\theta_*, \mu_*), 1} || \nu_{(\theta, \mu), 1}) + \lambda_2 \text{KL}(\nu_{(\theta_*, \mu_*), 2} || \nu_{(\theta, \mu), 2})}$$

$$= \log \frac{1}{2.4\delta} \cdot \frac{2}{\|\theta_* - \theta\|^2_{A(\widetilde{\lambda})} + \|\mu_* - \mu\|^2_{A(\widetilde{\lambda})}}.$$

where the equality follows by the same calculation as in the proof of Theorem 29. Take (θ, μ) to be $\theta = \theta_*$, $\mu = [0, 1 - \frac{\alpha}{1+2\alpha}]$ and note that $(\theta, \mu) \in \mathcal{C}_{\text{alt}}$ since with this choice of μ, arm z_2 is now safe. Now,

$$\|\mu_* - \mu\|^2_{A(\widetilde{\lambda})} = (\frac{\alpha}{1+2\alpha})^2 \cdot \frac{2\alpha}{1+2\alpha} = \frac{2\alpha^3}{(1+2\alpha)^3}.$$

This gives a lower bound of

$$\mathbb{E}[T] \geq \log \frac{1}{2.4\delta} \cdot \frac{2(1+2\alpha)^3}{2\alpha^3} \geq \log \frac{1}{2.4\delta} \cdot \frac{1}{\alpha^4}.$$

This lower bound is for best-arm identification ($\epsilon = 0$), but setting $\alpha \leftarrow 2\epsilon - a$ for a arbitrarily small, identifying an ϵ-optimal, ϵ-safe arm is equivalent to identifying the best arm, so this therefore holds as a lower bound on (ϵ, δ)-PAC algorithms.

The upper bound on the performance of BESIDE follows trivially since by setting $\lambda_1 = \lambda_2$, we can make the numerator in both terms of the complexity $\mathcal{O}(1)$, and the denominator of each term will be at least ϵ^2.

Proof for $\mathcal{I}^2$. Fix $\alpha \in (0, 0.1)$ and consider the following instance with $m = 1$:

$$\mathcal{X} = \{e_1, e_2\}, \quad \mathcal{Z} = \{z_1, z_2\}, \quad z_1 = [1/2 + \alpha^2/2, 0], \quad z_2 = [1/2, \alpha/2]$$
$$\theta_* = [1/2, 0], \quad \mu_* = [0, 0], \quad \gamma = 1.$$

On this instance, both z_1 and z_2 are safe, and z_1 is optimal.

The $\mathcal{XY}_{\text{safe}}$ design minimizes:

$$\max_{z \in \mathcal{Z}} \|z\|^2_{A(\lambda)^{-1}} = \max \left\{ \frac{1 + 2\alpha + \alpha^2}{4\lambda_1}, \frac{1}{4\lambda_1} + \frac{\alpha^2}{4\lambda_2} \right\}.$$

Some computation shows that, for α small, the optimal settings are $\lambda_1 = \mathcal{O}(1)$ and $\lambda_2 = \mathcal{O}(\alpha)$ (where here $\mathcal{O}(\cdot)$ hides terms that are $o(\alpha)$). Denote this allocation as $\widetilde{\lambda}$. Following the same argument as above, we have

$$\mathbb{E}[T] \geq \log \frac{1}{2.4\delta} \cdot \frac{2}{\|\theta_* - \theta\|^2_{A(\widetilde{\lambda})} + \|\mu_* - \mu\|^2_{A(\widetilde{\lambda})}}$$

for any $(\mu, \theta) \in \mathcal{C}_{\text{alt}}$. Let $\theta = [1/2, 2\alpha]$ and note that $(\mu_*, \theta) \in \mathcal{C}_{\text{alt}}$ since z_2 is now the optimal arm with this θ. We then have

$$\|\theta_* - \theta\|^2_{A(\widetilde{\lambda})} + \|\mu_* - \mu\|^2_{A(\widetilde{\lambda})} = \mathcal{O}(\alpha^3)$$

which gives a lower bound of

$$\mathbb{E}[T] \geq \Omega \left(\frac{1}{\alpha^3} \cdot \log \frac{1}{\delta} \right).$$

This lower bound holds for the best-arm identification problem, but setting $\alpha \leftarrow \sqrt{2\epsilon} - a$ for a arbitrarily small, finding an ϵ-optimal arm is equivalent to finding the best arm, so the lower bound applies in that setting as well.

To compute the sample complexity of BESIDE, we note that $\Delta_{\text{safe}}(z_1) = \Delta_{\text{safe}}(z_2) = 1$, so the first term in the complexity is negligible. We also have that $\Delta^{\widetilde{\epsilon}}(z_2) = \alpha^2/2 = \mathcal{O}(\epsilon)$ for $\widetilde{\epsilon} \leq 1$. Thus, the second term in the complexity scales as

$$\begin{aligned}\widetilde{\mathcal{O}}\left(\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z,z'\in\mathcal{Z}}\frac{\|z-z'\|^2_{A(\lambda)^{-1}}\cdot\log 1/\delta}{\epsilon^2}\right) &= \widetilde{\mathcal{O}}\left(\inf_{\lambda\in\triangle_{\mathcal{X}}}\frac{(\alpha^4/\lambda_1+\alpha^2/\lambda_2)\cdot\log 1/\delta}{\epsilon^2}\right)\\ &= \widetilde{\mathcal{O}}\left(\frac{\alpha^2\cdot\log 1/\delta}{\epsilon^2}\right)\\ &= \widetilde{\mathcal{O}}\left(\frac{\log 1/\delta}{\epsilon}\right).\end{aligned}$$

□

E.2 Robust Mean Estimation

In order to form estimates of $z^\top\theta_*$ and $z^\top\mu_{*,i}$, we will rely on the RIPS procedure proposed in (Camilleri et al., 2021a), instantiated with the robust Catoni estimator (Catoni, 2012).

Catoni Estimation. The robust Catoni mean estimator proposed in (Catoni, 2012) is defined as follows.

Definition 8 (Catoni Estimator)**.** *Consider real values $X_1,\dots,X_T$. Then the* robust Catoni mean estimator*, $\mathsf{cat}_\alpha[\{X_t\}_{t=1}^T]$, with parameter $\alpha>0$ is the unique root z of the function*

$$f_{\mathsf{cat}}(z;\{X_i\}_{i=1}^T,\alpha) := \sum_{t=1}^T \psi_{\mathsf{cat}}(\alpha(X_t-z)) \quad \textit{for} \quad \psi_{\mathsf{cat}}(y) := \begin{cases}\log(1+y+y^2) & y\geq 0\\ \log(1-y+y^2) & y<0\end{cases}.$$

The Catoni estimator satisfies the following guarantee.

Proposition 6. *Let $X_1,\dots,X_T$ be independent, identically distributed random variables with mean ζ and variance $\sigma^2<\infty$. Fix $\delta\in(0,1)$ and assume $T\geq 4\log(1/\delta)$. Then the Catoni estimator $\mathsf{cat}_\alpha[\{X_t\}_{t=1}^T]$ with parameter*

$$\alpha = \sqrt{\frac{2\log 1/\delta}{T\sigma^2}} \tag{E.1}$$

satisfies, with probability at least $1-2\delta$,

$$|\mathsf{cat}_\alpha[\{X_t\}_{t=1}^T]-\zeta| \leq \sqrt{\frac{8\sigma^2\log 1/\delta}{T}}.$$

Notably, the estimation error given by Proposition 6 scales only with the variance of the random variables, and not with their magnitude.

Robust Inverse Propensity Score (RIPS) Estimator. We apply the Catoni estimator with the RIPS estimator of (Camilleri et al., 2021a). In particular, consider running the following procedure.

We have the following guarantee on this procedure.

Algorithm E.1. Robust Inverse Propensity Score Estimation (RIPS)

1: **input:** samples $\{(x_t, r_t)\}_{t=1}^T$ for $x_t \sim \lambda$ and $r_t = \theta^\top x_t + w_t$, active set $\mathcal{Y}$, confidence δ
2: For each $y \in \mathcal{Y}$, set $W^y \leftarrow \mathsf{cat}_\alpha[\{y^\top A(\lambda)^{-1} x_t r_t\}_{t=1}^T]$, for α chosen as in (E.1) with $\delta \leftarrow \frac{\delta}{2|\mathcal{Y}|}$, and $A(\lambda) = \sum_{x \in \mathcal{X}} \lambda_x x x^\top$.
3: Set

$$\widehat{\theta} = \mathrm{argmin}_\theta \max_{y \in \mathcal{Y}} \frac{|\theta^\top y - W^y|}{\|y\|_{A(\lambda)^{-1}}}.$$

4: **return** $\widehat{\theta}$

Proposition 7 (Theorem 1 of (Camilleri et al., 2021a)). *If $T \geq 4 \log \frac{2|\mathcal{Z}|}{\delta}$, then with probability at least $1 - \delta$, for all $z \in \mathcal{Z}$, the RIPS estimator of Algorithm E.1 returns an estimate $\widehat{\theta}$ which satisfies:*

$$|y^\top(\widehat{\theta} - \theta_*)| \leq \|y\|_{A(\lambda)^{-1}} \cdot \sqrt{\frac{8 \log(2|\mathcal{Z}|/\delta)}{T}}.$$

The use of the RIPS estimator allows us to avoid sophisticated rounding procedures often found in the linear bandit literature. Note that the RIPS estimator can be computed in time scaling polynomially in $|\mathcal{Y}|$, d, and T.

E.3 RAGE$^\epsilon$

A note on constants. Throughout our algorithm definitions, in both this section and the following, we use generic constants rather than precise numerical settings, and carry these generic constants through our proofs. At various points in the proofs, we require that these constants satisfy certain constraints. The following result shows that there exist suitable settings for all constants such that these constraints are satisfied.

Lemma 59. *There exist settings of $c_a, c_b, c_d, c_e, c_f, c_e, c_g, c_1, c_2, c_3, c_4, c_\Delta$ and c_0 such that Equations* (E.6)*,* (E.9)*,* (E.10)*,* (E.11)*,* (E.12)*,* (E.13)*,* (E.14)*,* (E.2)*,* (E.3)*, and* (E.4) *are satisfied, and*

$$\frac{c_3(1 + c_g)}{1 - c_3} \leq 0.2, \quad c_g \leq 0.2, \quad c_0 \geq 0.0001.$$

Proof. First, note that in addition to the conditions listed above, we must also have

$$c_1 \leq c_f, \quad 3(c_d + c_e) \leq c_2.$$

Furthermore, by Lemma 70, it suffices to always take $c_\Delta = 3c_d + 3c_e - c_g$. Direct computation then shows that the following settings suffice, up to machine precision:

$$\begin{aligned} c_1 &= 0.05978841810030329 \\ c_2 &= 0.0600087370242953 \\ c_3 &= 0.1 \\ c_4 &= 0.1 \end{aligned}$$

$$
\begin{aligned}
c_a &= 0.0013004532984432395 \\
c_b &= 0.41043329378840077 \\
c_d &= 0.01 \\
c_e &= 0.01 \\
c_c &= 0.0014065949472697806 \\
c_g &= 0.178 \\
c_f &= c_1
\end{aligned}
$$

Given these settings, we can bound

$$\frac{c_3(1+c_g)}{1-c_3} \leq 0.5.$$

□

E.3.1 Preliminaries

Assumptions and Definitions. For all $y \in \mathcal{Y}$, $\widehat{\Delta}_{\text{safe}}(y) \geq -c_\Delta \epsilon$. We will also assume that $\mathcal{Y} \subseteq \mathcal{X}$. We define

$$y_\star = \text{argmin}_{y \in \mathcal{Y}} y^\top \theta_*$$

and

$$\Delta(z) = \theta_*^\top (z - y_\star).$$

We will take $\gamma = 0$, so we set $A(\lambda) = \sum_{x \in \mathcal{X}} \lambda_x x x^\top$.

E.3.2 Algorithm and Main Results

At a high-level, RAGE$^\epsilon$ attempts to estimate the difference between the performance of each $z \in \mathcal{Z}$ and the *best* $y \in \mathcal{Y}$. The safety gap estimate, $\widehat{\Delta}_{\text{safe}}(z)$, acts as a regularizer: if $\widehat{\Delta}_{\text{safe}}(z) < 0$, then we do not seek to estimate the gap of z with as high accuracy, since we can already eliminate it by showing it is unsafe. The proof in this section follow closely the proof given in Section 6.4.4 of (Jamieson and Jain, 2022).

Theorem 30. *With probability at least $1-\delta$,* RAGE$^\epsilon$ *will terminate after collecting at most*

$$C \cdot \sum_{\ell=1}^{\lceil \log 2/c_f\epsilon \rceil} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} \frac{\|z - y_\star\|^2_{A(\lambda)^{-1}}}{(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\Delta(z)) + \epsilon_\ell)^2} \cdot \log(\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta}) + 4\lceil \log \tfrac{2}{c_f\epsilon} \rceil \log(\tfrac{4|\mathcal{Z}|^2 \lceil \log \frac{2}{c_f\epsilon} \rceil}{\delta})$$

samples, for a universal constant C, and will output estimates of the gaps $\widehat{\Delta}(z)$ such that, for all $z \in \mathcal{Z}$,

$$|\widehat{\Delta}(z) - \Delta(z)| \leq c_f \left(\epsilon + \mathfrak{p}(\Delta(z)) + \mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) \right).$$

Algorithm E.2. RAGE$^\epsilon$

input: active set $\mathcal{Z}$, optimal set $\mathcal{Y}$, tolerance ϵ, confidence δ, safety gap estimate $\{\widehat{\Delta}_{\text{safe}}(z)\}_{z\in\mathcal{Z}}$
Choose $\widehat{y}_0$ arbitrarily from $\mathcal{Y}$, set $\widehat{\Delta}^0(z) \leftarrow 0$ for all $z \in \mathcal{Z}$
for $\ell = 1, 2, \ldots, \lceil \log(2/c_f\epsilon)\rceil$ **do**
    $\epsilon_\ell \leftarrow \frac{2}{c_f} \cdot 2^{-\ell}$
    Let τ_ℓ be the minimal value of $\tau = 2^j \geq 4\log\frac{4|\mathcal{Z}|^2\ell^2}{\delta}$ such that the objective to the following is no greater than $c_c\epsilon_\ell$, and λ_ℓ the corresponding optimal distribution

$$\inf_{\lambda\in\triangle_\mathcal{X}} \max_{z\in\mathcal{Z}} -c_a(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell) + \sqrt{\frac{\|z-\widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau}}.$$

    Sample $x_t \sim \lambda_\ell$, collect observations $\{(x_t, r_t, s_{t,1}, \ldots, s_{t,m})\}_{t=1}^{\tau_\ell}$
    $\mathcal{W} \leftarrow \{z - z' \,:\, z, z' \in \mathcal{Z}\}$
    $\widehat{\theta}^\ell \leftarrow \mathsf{RIPS}(\{(x_t, r_t)\}_{t=1}^{\tau_\ell}, \mathcal{W}, \frac{\delta}{2\ell^2})$
    Set

$$\widehat{y}_\ell \leftarrow \text{argmin}_{y\in\mathcal{Y}} y^\top\widehat{\theta}^\ell + 8\sqrt{\frac{\|y-\widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}}$$

$$\widehat{\Delta}^\ell(y) \leftarrow (y-\widehat{y}_\ell)^\top\widehat{\theta}^\ell + \sqrt{\frac{\|y-\widehat{y}_\ell\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}}$$

return $\{\widehat{\Delta}^\ell(z)\}_{z\in\mathcal{Z}}$

E.3.3 Estimating the Gaps

Lemma 60. *Let $\mathcal{E}_{\text{RAGE}^\epsilon}$ denote the event that for all ℓ and all $z, z' \in \mathcal{X}$, we have:*

$$|(\widehat{\theta}^\ell - \theta_*)^\top(z-z')| \leq \sqrt{8\|z-z'\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log\frac{4|\mathcal{Z}|^2\ell^2}{\delta}}{\tau_\ell}}.$$

Then $\Pr[\mathcal{E}_{\text{RAGE}^\epsilon}] \geq 1-\delta$.

Proof. Since $\tau_\ell \geq 4\log\frac{4|\mathcal{Z}|^2\ell^2}{\delta}$, we can apply Proposition 7 to get that, with probability at least $1-\delta/2\ell^2$, for all $w \in \mathcal{W}$,

$$|(\widehat{\theta}^\ell - \theta_*)^\top w| \leq \sqrt{8\|w\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log\frac{4|\mathcal{Z}|^2\ell^2}{\delta}}{\tau_\ell}}.$$

The result then follows by a union bound since

$$\sum_{\ell=1}^{\infty} \frac{\delta}{2\ell^2} = \frac{\pi^2}{12}\delta \leq \delta.$$

□

Lemma 61. *On $\mathcal{E}_{\text{RAGE}^\epsilon}$, for all $z \in \mathcal{Z}$ and all ℓ,*

$$|\widehat{\Delta}^\ell(z) - \theta_*^\top(z - \widehat{y}_\ell)| \le 8c_a\left(\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(\widehat{y}_\ell))\right) + 8(c_c + c_a + 2c_a c_\Delta)\epsilon_\ell.$$

Proof. By construction, we have that

$$\max_{z\in\mathcal{Z}} -c_a(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell) + \sqrt{\frac{\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \le c_c\epsilon_\ell.$$

This implies that, for all $z \in \mathcal{Z}$:

$$\sqrt{\frac{\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \le c_a(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z))) + (c_c + c_a)\epsilon_\ell.$$

On $\mathcal{E}_{\text{RAGE}^\epsilon}$, we have

$$\begin{aligned}
|\widehat{\Delta}_\ell(z) - \theta_*^\top(z - \widehat{y}_\ell)| &\le \sqrt{8\|z - \widehat{y}_\ell\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&\le \sqrt{16\|\widehat{y}_{\ell-1} - \widehat{y}_\ell\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} + \sqrt{16\|\widehat{y}_{\ell-1} - \widehat{y}_\ell\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&\le 8c_a\left(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(\widehat{y}_\ell)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(\widehat{y}_\ell))\right) + 8(c_c + c_a)\epsilon_\ell.
\end{aligned}$$

By construction we have that $\widehat{\Delta}_{\text{safe}}(\widehat{y}_\ell) \ge -c_\Delta\epsilon \ge -2c_\Delta\epsilon_\ell$, so $\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(\widehat{y}_\ell)) \le 2c_\Delta\epsilon_\ell$, which proves the result. □

Lemma 62. *On $\mathcal{E}_{\text{RAGE}^\epsilon}$ and the event that $\widehat{\Delta}^{\ell-1}(y_\star) \le c_b\epsilon_\ell$, we have*

$$\Delta(\widehat{y}_\ell) \le 6(c_c + c_a(1 + c_b + 2c_\Delta))\epsilon_\ell.$$

Proof. By the definition of $\mathcal{E}_{\text{RAGE}^\epsilon}$ and $\widehat{y}_\ell$, we can bound

$$\begin{aligned}
\theta_*^\top(\widehat{y}_\ell - \widehat{y}_{\ell-1}) &\le (\widehat{\theta}^\ell)^\top(\widehat{y}_\ell - \widehat{y}_{\ell-1}) + \sqrt{8\|\widehat{y}_\ell - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&= \min_{y\in\mathcal{Y}} (\widehat{\theta}^\ell)^\top(y - \widehat{y}_{\ell-1}) + \sqrt{8\|y - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&\le (\widehat{\theta}^\ell)^\top(y_\star - \widehat{y}_{\ell-1}) + \sqrt{8\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&\le \theta_*^\top(y_\star - \widehat{y}_{\ell-1}) + 2\sqrt{8\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}}.
\end{aligned}$$

By the definition of τ_ℓ and λ_ℓ, we have

$$c_c\epsilon_\ell \ge \max_{z\in\mathcal{Z}} -c_a(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell) + \sqrt{\frac{\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}}$$

$$
\begin{aligned}
&\geq -c_a(\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(y_\star)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(y_\star)) + \epsilon_\ell) + \sqrt{\frac{\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&\overset{(a)}{\geq} -c_a(\mathfrak{p}(\widehat{\Delta}^{\ell-1}(y_\star)) + (1+2c_\Delta)\epsilon_\ell) + \sqrt{\frac{\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \\
&\overset{(b)}{\geq} -c_a(1+c_b+2c_\Delta)\epsilon_\ell + \sqrt{\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}}
\end{aligned}
$$

where (a) uses that $\widehat{\Delta}_{\mathrm{safe}}(y_\star) \geq -c_\Delta\epsilon \geq -2c_\Delta\epsilon_\ell$, by definition, and (b) follows by our assumption on $\widehat{\Delta}^{\ell-1}(y_\star)$. This implies that

$$
\sqrt{\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4|\mathcal{Z}|^2\ell^2}{\delta})}{\tau_\ell}} \leq (c_c + c_a(1+c_b+2c_\Delta))\epsilon_\ell.
$$

Combining this with the above we have that

$$
\theta_*^\top(\widehat{y}_\ell - \widehat{y}_{\ell-1}) \leq \theta_*^\top(y_\star - \widehat{y}_{\ell-1}) + 6(c_c + c_a(1+c_b+2c_\Delta))\epsilon_\ell.
$$

Rearranging this proves the result. □

Lemma 63. *For all $z \in \mathcal{Z}$ and all ℓ, on the event $\mathcal{E}_{\mathrm{RAGE}^\epsilon}$,*

$$
|\widehat{\Delta}^\ell(z) - \Delta(z)| \leq c_f\left(\epsilon_\ell + \mathfrak{p}(\Delta(z)) + \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z))\right).
$$

Proof. We prove this by induction. Assume that at $\ell - 1$, for all $z \in \mathcal{Z}$,

$$
|\widehat{\Delta}^{\ell-1}(z) - \Delta(z)| \leq c_f\left(\epsilon_{\ell-1} + \mathfrak{p}(\Delta(z)) + \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z))\right).
$$

On $\mathcal{E}_{\mathrm{RAGE}^\epsilon}$ and by Lemma 61 we can bound

$$
\begin{aligned}
|\widehat{\Delta}^\ell(z) - \Delta(z)| &= |\widehat{\Delta}^\ell(z) - (R(z) - R(\widehat{y}_\ell) + R(\widehat{y}_\ell) - R(y_\star))| \\
&\leq |\widehat{\Delta}^\ell(z) - (R(z) - R(\widehat{y}_\ell))| + \Delta(\widehat{y}_\ell) \\
&\leq 8c_a\left(\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(\widehat{y}_\ell))\right) + 8(c_c + c_a + 2c_ac_\Delta)\epsilon_\ell + \Delta(\widehat{y}_\ell).
\end{aligned}
$$

By the inductive hypothesis, we can bound

$$
\begin{aligned}
\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) &\leq (1+c_f)\mathfrak{p}(\Delta(z)) + c_f\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + c_f\epsilon_{\ell-1} \\
\mathfrak{p}(\widehat{\Delta}^{\ell-1}(\widehat{y}_\ell)) &\leq (1+c_f)\mathfrak{p}(\Delta(\widehat{y}_\ell)) + c_f\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(\widehat{y}_\ell)) + c_f\epsilon_{\ell-1}.
\end{aligned}
$$

By construction $\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(\widehat{y}_\ell)) \geq -c_\Delta\epsilon \geq -2c_\Delta\epsilon_\ell$, so

$$
\begin{aligned}
|\widehat{\Delta}^\ell(z) - \Delta(z)| \leq\ & 8c_a(1+c_f)\mathfrak{p}(\Delta(z)) + 8c_a(1+c_f)\mathfrak{p}(\widehat{\Delta}_{\mathrm{safe}}(z)) + (8c_a(1+c_f)+1)\Delta(\widehat{y}_\ell) \\
&+ 8(c_ac_f(1+c_\Delta) + c_c + c_a + 2c_ac_\Delta)\epsilon_\ell.
\end{aligned}
$$

It remains to bound $\Delta(\widehat{y}_\ell) = R(\widehat{y}_\ell) - R(y_\star)$. On the inductive hypothesis, we have that

$$|\widehat{\Delta}^{\ell-1}(y_\star) - \Delta(y_\star)| \le c_f \left(\epsilon_{\ell-1} + \mathfrak{p}(\Delta(y_\star)) + \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(y_\star))\right).$$

By definition, $\Delta(y_\star) = 0$ and $\widehat{\Delta}_{\mathrm{safe}}(y_\star) \ge -c_\Delta \epsilon \ge -2c_\Delta \epsilon_\ell$, which implies that $\widehat{\Delta}^{\ell-1}(y_\star) \le 2c_f(1+c_\Delta)\epsilon_\ell$. It follows that the conditions of Lemma 62 are met as long as

$$2c_f(1 + c_\Delta) \le c_b, \tag{E.2}$$

so we can bound $\Delta(\widehat{y}_\ell) \le 6(c_c + c_a(1 + c_b + 2c_\Delta))\epsilon_\ell$. Thus,

$$|\widehat{\Delta}^{\ell}(z) - \Delta(z)| \le 8c_a(1 + c_f)\mathfrak{p}(\Delta(z)) + 8c_a(1 + c_f)\mathfrak{p}(\widehat{\Delta}_{\mathrm{safe}}(z)) + 8(c_a c_f(1 + c_\Delta) + c_c + c_a + 2c_a c_\Delta)\epsilon_\ell \\ + (8c_a(1 + c_f) + 1)(6(c_c + c_a(1 + c_b + 2c_\Delta)))\epsilon_\ell.$$

which proves the inductive hypothesis as long as

$$8(c_a c_f(1 + c_\Delta) + c_c + c_a + 2c_a c_\Delta) + (8c_a(1 + c_f) + 1)(6(c_c + c_a(1 + c_b + 2c_\Delta))) \le c_f \\ 8c_a(1 + c_f) \le c_f \tag{E.3}$$

For the base case, we need to show that

$$|\widehat{\Delta}^{0}(z) - \Delta(z)| \le c_f \left(\epsilon_0 + \mathfrak{p}(\Delta(z)) + \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z))\right).$$

By construction $\widehat{\Delta}^0(z) = 0$ for all z, and $\mathfrak{p}(\Delta(z)) \ge 0$, $\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) \ge 0$. Thus, it suffices to show $|\Delta(z)| \le c_f \epsilon_0$. However, by construction $|\Delta(z)| \le 1$, and $c_f \epsilon_0 = 1$, which proves the base case. □

E.3.4 Bounding the Sample Complexity

Lemma 64. *On the event $\mathcal{E}_{\mathrm{RAGE}^\epsilon}$, RAGE^ϵ will terminate after collecting at most*

$$C \cdot \sum_{\ell=1}^{\lceil \log(2/c_f\epsilon)\rceil} \inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} \frac{\|z - y_\star\|^2_{A(\lambda)^{-1}}}{(\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + \mathfrak{p}(\Delta(z)) + \epsilon_\ell)^2} \cdot \log(\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta}) + 8\lceil \log \tfrac{2}{c_f\epsilon}\rceil \log(\tfrac{4|\mathcal{Z}|^2\lceil \log \frac{2}{c_f\epsilon}\rceil^2}{\delta})$$

samples, for a universal constant C.

Proof. If, for all $z \in \mathcal{Z}$,

$$\tau \ge \frac{\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}}}{(c_a(\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell) + c_c\epsilon_\ell)^2} \cdot \log(\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta})$$

we will have that the objective on Line 5 of RAGE^ϵ is less than $c_c\epsilon_\ell$. Since we can take the best-case $\lambda \in \triangle_\mathcal{X}$, and since we have that τ_ℓ will be at most a factor of 2 from the optimal τ, it follows that

$$\begin{aligned}\tau_\ell &\le \inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} \frac{2\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}}}{(c_a(\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell) + c_c\epsilon_\ell)^2} \cdot \log(\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta}) \vee 8\log\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta} \\ &\le \inf_{\lambda \in \triangle_\mathcal{X}} \max_{z \in \mathcal{Z}} \frac{2\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}}}{(c_a(\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell) + c_c\epsilon_\ell)^2} \cdot \log(\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta}) + 8\log\tfrac{4|\mathcal{Z}|^2\ell^2}{\delta}\end{aligned}$$

where the additional $8 \log \frac{4|\mathcal{Z}|^2\ell^2}{\delta}$ factor arises since we always require $\tau_\ell \geq 4 \log \frac{4|\mathcal{Z}|^2\ell^2}{\delta}$.

We can upper bound

$$\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}} \leq 2\|z - y_\star\|^2_{A(\lambda)^{-1}} + 2\|y_\star - \widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}}.$$

By construction, $\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(\widehat{y}_{\ell-1})) \leq 2c_\Delta\epsilon_\ell$, so for any z, $\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(\widehat{y}_{\ell-1})) - 2c_\Delta\epsilon_\ell \leq \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z))$. Furthermore, by definition,

$$\widehat{\Delta}^{\ell-1}(\widehat{y}_{\ell-1}) = 0$$

so $\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) \geq \mathfrak{p}(\widehat{\Delta}^{\ell-1}(\widehat{y}_{\ell-1}))$. Thus,

$$\begin{aligned}
&\inf_{\lambda\in\triangle_{\mathcal{X}}} \max_{z\in\mathcal{Z}} \frac{\|z - \widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}}}{(c_a\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + c_a\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + (c_a + c_c)\epsilon_\ell)^2} \\
&\leq \inf_{\lambda\in\triangle_{\mathcal{X}}} \max_{z\in\mathcal{Z}} \frac{2\|z - y_\star\|^2_{A(\lambda)^{-1}}}{(c_a\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + c_a\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + (c_a + c_c)\epsilon_\ell)^2} \\
&\quad + \frac{2\|\widehat{y}_{\ell-1} - y_\star\|^2_{A(\lambda)^{-1}}}{(c_a\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(\widehat{y}_{\ell-1})) + c_a\mathfrak{p}(\widehat{\Delta}^{\ell-1}(\widehat{y}_{\ell-1})) + (c_a + c_c - 2c_ac_\Delta)\epsilon_\ell)^2} \\
&\leq \inf_{\lambda\in\triangle_{\mathcal{X}}} \max_{z\in\mathcal{Z}} \frac{4\|z - y_\star\|^2_{A(\lambda)^{-1}}}{(c_a\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + c_a\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + (c_a + c_c - 2c_ac_\Delta)\epsilon_\ell)^2}.
\end{aligned}$$

By Lemma 63, we can lower bound

$$\widehat{\Delta}^{\ell-1}(z) \geq \Delta(z) - c_f(\epsilon_\ell + \mathfrak{p}(\Delta(z)) + \mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)))$$

so

$$\begin{aligned}
&c_a\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + c_a\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + (c_a + c_c - 2c_ac_\Delta)\epsilon_\ell \\
&\quad \geq c_a(1 - c_f)\mathfrak{p}(-\widehat{\Delta}_{\mathrm{safe}}(z)) + c_a(1 - c_f)\mathfrak{p}(\Delta(z)) + (c_a + c_c - 2c_ac_\Delta - c_ac_f)\epsilon_\ell.
\end{aligned}$$

The result follows by combining these inequalities and as long as

$$c_a(1 - c_f) \geq c_0, \qquad c_a + c_c - 2c_ac_\Delta - c_ac_f \geq c_0. \tag{E.4}$$

□

Proof of Theorem 30. Theorem 30 follows directly from Lemma 64 and Lemma 63 since, by Lemma 60, $\mathcal{E}_{\textsc{Rage}^\epsilon}$ holds with probability at least $1 - \delta$. □

E.4 Safe Best-Arm Identification

E.4.1 Preliminaries

In general we want to consider multiple safety constraints, and let m denote the number of constraints. In such settings, we will denote $\Delta^i_{\mathrm{safe}}(z)$ the safety gap for safety constraint i.

Define

$$\widetilde{\Delta}^\ell(z) := \theta_*^\top z - \min_{y\in\mathcal{Y}_\ell} \theta_*^\top y.$$

E.4.2 Algorithm and Main Result

Theorem 31 (Full version of Theorem 16)**.** *With probability at least* $1-2\delta$*, Algorithm 6.1 returns an arm* $\widehat{z}$ *such that*

$$\widehat{z}^\top \theta_* \geq (z_*)^\top \theta_* - \epsilon, \quad \Delta_{\mathrm{safe}}(\widehat{z}) \geq -\epsilon \tag{E.5}$$

and terminates after collecting at most

$$C \cdot \sum_{\ell=1}^{\iota_\epsilon} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} \frac{\|z\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\left(\min_j |\Delta^j_{\mathrm{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\mathrm{safe}}(z)) + \mathfrak{p}(\Delta^{\epsilon_{\ell-1}}(z)) + \epsilon_\ell\right)^2}$$
$$+ C \log \frac{1}{\epsilon} \cdot \sum_{\ell=1}^{\iota_\epsilon} \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} \frac{\|z - z_*\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{8|\mathcal{Z}|^2 \log^4(1/\epsilon)}{\delta})}{(\max_j \mathfrak{p}(-\Delta^j_{\mathrm{safe}}(z)) + \mathfrak{p}(\Delta^{\epsilon_\ell}(z)) + \epsilon_\ell)^2} + C_0$$

samples for a universal constant C*,* $C_0 = \mathrm{poly}\log(\frac{1}{\epsilon}, |\mathcal{Z}|) \cdot \log \frac{1}{\delta}$.

E.4.3 Estimating the Safety Value

Lemma 65. *Let* $\mathcal{E}_{\mathrm{safe}}$ *denote the event that, for all* ℓ*,* $z \in \mathcal{Z}$*,* $i \in [m]$*:*

$$|z^\top(\widehat{\mu}^{i,\ell} - \mu^i_*)| \leq \sqrt{8\|z\|^2_{A(\lambda_\ell)^{-1}} \cdot \frac{\log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau_\ell}}.$$

Then $\Pr[\mathcal{E}_{\mathrm{safe}}] \geq 1-\delta$.

Proof. This follows directly from Proposition 7 and a union bound, as in Lemma 60. □

Lemma 66. *On* $\mathcal{E}_{\mathrm{safe}}$*, for all* $z \in \mathcal{Z}$*,* $i \in [m]$*, and all* ℓ*,*

$$|\widehat{\Delta}^{i,\ell}_{\mathrm{safe}}(z) - \Delta^i_{\mathrm{safe}}(z)| \leq 3c_d \left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z))\right) + 3(c_d + c_e)\epsilon_\ell.$$

Proof. By construction, we have that

$$\max_{z \in \mathcal{Z}} -c_d \left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell\right) + \sqrt{\frac{\|z\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau_\ell}} \leq c_e\epsilon_\ell.$$

This implies that, for all $z \in \mathcal{Z}$,

$$\sqrt{\frac{\|z\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau_\ell}} \leq \min_j c_d|\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)| + \max_j c_d\mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)) + c_d\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + (c_d + c_e)\epsilon_\ell.$$

On $\mathcal{E}_{\mathrm{safe}}$, we have

$$|\widehat{\Delta}^{i,\ell}_{\mathrm{safe}}(z) - \Delta^i_{\mathrm{safe}}(z)| \leq \sqrt{8\frac{\|z\|^2_{A(\lambda_\ell)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau_\ell}}$$
$$\leq \min_j 3c_d|\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)| + \max_j 3c_d\mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\mathrm{safe}}(z)) + 3c_d\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + 3(c_d + c_e)\epsilon_\ell$$

which proves the result. □

E.4.4 Tying Together Safety Estimation with Optimality Estimation

Definition 9 (Optimality Good Event)**.** *Let $\mathcal{E}^{\ell}_{\text{RAGE}^\epsilon}$ denote the success event of* RAGE$^\epsilon$ *when called at the ℓth epoch, and $\mathcal{E}_{\text{RAGE}^\epsilon} := \cup_\ell \mathcal{E}^{\ell}_{\text{RAGE}^\epsilon}$.*

Lemma 67. *On the event $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$, we have that:*

1. *For all $\ell \le \iota_\epsilon$, $y \in \mathcal{Y}_\ell$, and $i \in [m]$, $y^\top \mu_{*,i} \le \gamma$.*
2. *For all ℓ and $z \in \mathcal{Z}$, $\widetilde{\Delta}^{\ell-1}(z) \le \widetilde{\Delta}^{\ell}(z)$.*

Proof. By Lemma 66, we have that

$$\widehat{\Delta}^{i,\ell}_{\text{safe}}(z) - 3c_d\left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z))\right) - 3(c_d+c_e)\epsilon_\ell \le \Delta^i_{\text{safe}}(z).$$

Thus, if the inclusion condition of $\mathcal{Y}_\ell$ is met, it must be the case that $\Delta^i_{\text{safe}}(z) \ge 0$ for all i.

The second conclusion follows directly since $\mathcal{Y}_{\ell-1} \subseteq \mathcal{Y}_\ell$. □

Lemma 68 (Key Estimation Error Bound)**.** *On the event $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$, for all $z \in \mathcal{Z}$, ℓ, and i, we have*

$$|\widehat{\Delta}^{\ell}(z) - \widetilde{\Delta}^{\ell}(z)| \le c_3\left(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^{\ell}(z)) + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z))\right)$$
$$|\widehat{\Delta}^{i,\ell}_{\text{safe}}(z) - \Delta^i_{\text{safe}}(z)| \le c_4\left(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^{\ell}(z)) + \min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z))\right).$$

Proof. We prove this by induction. Assume that the above inequalities hold at epoch $\ell-1$. On $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$, by Lemma 63 and Lemma 66, we have

$$|\widehat{\Delta}^{\ell}(z) - \widetilde{\Delta}^{\ell}(z)| \le c_1(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^{\ell}(z)) + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)))$$
$$|\widehat{\Delta}^{i,\ell}_{\text{safe}}(z) - \Delta^i_{\text{safe}}(z)| \le c_2(\epsilon_\ell + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z))).$$

By the inductive hypothesis, we can bound

$$\begin{aligned}\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) &\le \mathfrak{p}\left(\widetilde{\Delta}^{\ell-1}(z) + c_3(\epsilon_{\ell-1} + \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)))\right)\\ &\le (1+c_3)\mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) + c_3\epsilon_{\ell-1} + \max_j c_3\mathfrak{p}(-\Delta^j_{\text{safe}}(z))\\ &\le (1+c_3)\mathfrak{p}(\widetilde{\Delta}^{\ell}(z)) + 2c_3\epsilon_\ell + \max_j c_3\mathfrak{p}(-\Delta^j_{\text{safe}}(z))\end{aligned}$$

where the last inequality follows since, by Lemma 67, $\widetilde{\Delta}^{\ell-1}(z) \le \widetilde{\Delta}^{\ell}(z)$.

Furthermore, again applying the inductive hypothesis,

$$\begin{aligned}\widehat{\Delta}^{i,\ell-1}_{\text{safe}}(z) &\le \Delta^i_{\text{safe}}(z) + c_4(\epsilon_{\ell-1} + \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) + \min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)))\\ &\le \Delta^i_{\text{safe}}(z) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^{\ell}(z)) + \min_j c_4|\Delta^j_{\text{safe}}(z)| + \max_j c_4\mathfrak{p}(-\Delta^j_{\text{safe}}(z))\\ &\le \Delta^i_{\text{safe}}(z) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^{\ell}(z)) + c_4|\Delta^i_{\text{safe}}(z)| + \max_j c_4\mathfrak{p}(-\Delta^j_{\text{safe}}(z)).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathfrak{p}(-\widehat{\Delta}^{i,\ell-1}_{\text{safe}}(z)) &\le \mathfrak{p}\left(-\Delta^i_{\text{safe}}(z) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \min_j c_4|\Delta^j_{\text{safe}}(z)| + \max_j c_4\mathfrak{p}(-\Delta^j_{\text{safe}}(z))\right)\\ &\le \mathfrak{p}\left(-\Delta^i_{\text{safe}}(z) + \min_j c_4|\Delta^j_{\text{safe}}(z)|\right) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \max_j c_4\mathfrak{p}(-\Delta^j_{\text{safe}}(z))\\ &\le \mathfrak{p}\left(-\Delta^i_{\text{safe}}(z) + c_4|\Delta^i_{\text{safe}}(z)|\right) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \max_j c_4\mathfrak{p}(-\Delta^j_{\text{safe}}(z)).\end{aligned}$$

Note that if $\Delta^i_{\text{safe}}(z) \le 0$, then

$$\mathfrak{p}(-\Delta^i_{\text{safe}}(z) + c_4|\Delta^i_{\text{safe}}(z)|) = \mathfrak{p}(-\Delta^i_{\text{safe}}(z) - c_4\Delta^i_{\text{safe}}(z)) = (1+c_4)\mathfrak{p}(-\Delta^i_{\text{safe}}(z))$$

and if $\Delta^i_{\text{safe}}(z) > 0$, then for $c_4 < 1$, $-\Delta^i_{\text{safe}}(z) + c_4|\Delta^i_{\text{safe}}(z)| \le 0$, so

$$\mathfrak{p}(-\Delta^i_{\text{safe}}(z) + c_4|\Delta^i_{\text{safe}}(z)|) = 0 = (1+c_4)\mathfrak{p}(-\Delta^i_{\text{safe}}(z)).$$

Thus,

$$\mathfrak{p}(-\widehat{\Delta}^{i,\ell-1}_{\text{safe}}(z)) \le (1+c_4)\mathfrak{p}(-\Delta^i_{\text{safe}}(z)) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \max_j c_4\mathfrak{p}(-\Delta^j_{\text{safe}}(z)).$$

Combining these inequalities, it follows that

$$\begin{aligned}|\widehat{\Delta}^{i,\ell}_{\text{safe}}(z) - \Delta^i_{\text{safe}}(z)| &\le c_2\left(\epsilon_\ell + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z))\right)\\ &\le c_2(1+2c_3+4c_4)\epsilon_\ell + c_2(1+c_3+2c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z))\\ &\quad + c_2(1+c_3+3c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + c_2(1+c_4)\min_j |\Delta^j_{\text{safe}}(z)|\end{aligned}$$

and

$$\begin{aligned}|\widehat{\Delta}^\ell(z) - \widetilde{\Delta}^\ell(z)| &\le c_1(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)))\\ &\le c_1(1+2c_4)\epsilon_\ell + c_1(1+c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + c_1(1+2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)).\end{aligned}$$

This proves the inductive hypothesis, as long as

$$c_1(1+2c_4) \le c_3, \quad c_2(1+2c_3+4c_4) \le c_4. \tag{E.6}$$

For the base case, we need to show that

$$\begin{aligned}|\widehat{\Delta}^0(z) - \widetilde{\Delta}^0(z)| &\le c_3(\epsilon_0 + \mathfrak{p}(\widetilde{\Delta}^0(z)) + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)))\\ |\widehat{\Delta}^0_{\text{safe}}(z) - \Delta_{\text{safe}}(z)| &\le c_4(\epsilon_0 + \mathfrak{p}(\widetilde{\Delta}^0(z)) + \min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z))).\end{aligned}$$

By construction, $\widehat{\Delta}^0(z) = \widehat{\Delta}^0_{\text{safe}}(z) = 0$. Thus, it suffices to show $|\widetilde{\Delta}^0(z)| \le c_3\epsilon_0$ and $|\Delta_{\text{safe}}(z)| \le c_4\epsilon_0$. However, both of these are true by our choice of ϵ_0. □

Lemma 69. *On the event $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$, for all $z \in \mathcal{Z}$ and all ℓ, we will have*

$$\widetilde{\Delta}^\ell(z) \geq \Delta^{\epsilon_\ell}(z) \quad \textit{where} \quad \Delta^{\epsilon_\ell}(z) = \max_{y \in \mathcal{Z} \,:\, \epsilon_\ell \leq \min_i \Delta^i_{\text{safe}}(y)} y^\top \theta_* - z^\top \theta_*.$$

Proof. By definition, we will have $z \in \mathcal{Y}_\ell$ if

$$8c_d \left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) \right) + 8(c_d + c_e)\epsilon_\ell \leq \widehat{\Delta}^{i,\ell}_{\text{safe}}(z).$$

The following claim allows us to obtain a sufficient condition to guarantee $z \in \mathcal{Y}_\ell$.

Claim 1. *On the event $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$,*

$$\begin{aligned}
&\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) \\
&\quad \leq 2(c_3 + 2c_4)\epsilon_\ell + (1 + c_3 + 2c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + (1 + 2c_4) \min_j |\Delta^j_{\text{safe}}(z)| + (1 + c_3 + 2c_4) \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)).
\end{aligned}$$

Proof of Claim 1. By Lemma 67 and Lemma 68, we can bound

$$\begin{aligned}
\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| &\leq (1 + c_4) \min_j |\Delta^j_{\text{safe}}(z)| + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + c_4 \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) \\
\max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) &\leq (1 + c_4) \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + 2c_4\epsilon_\ell + c_4\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + c_4 \min_j |\Delta^j_{\text{safe}}(z)| \\
\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) &\leq (1 + c_3)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + 2c_3\epsilon_\ell + c_3 \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)).
\end{aligned}$$

The claim follows by summing these upper bounds. □

Thus, by Claim 1, we can bound

$$\begin{aligned}
&3c_d \left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) \right) + 3(c_d + c_e)\epsilon_\ell \\
&\quad \leq 3(c_d + c_e + 2c_dc_3 + 4c_dc_4)\epsilon_\ell + 3c_d(1 + c_3 + 2c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) \\
&\quad + 3c_d(1 + 2c_4) \min_j |\Delta^j_{\text{safe}}(z)| + 3c_d(1 + c_3 + 2c_4) \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)).
\end{aligned}$$

Furthermore, by Lemma 68,

$$\Delta^i_{\text{safe}}(z) - c_4 \left(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) \right) \leq \widehat{\Delta}^{i,\ell}_{\text{safe}}(z)$$

It follows that a sufficient condition for $z \in \mathcal{Y}_\ell$ is

$$\begin{aligned}
&(3c_d + 3c_e + 6c_dc_3 + 12c_dc_4 + c_4) \left(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) \right) \\
&\quad \leq \Delta^i_{\text{safe}}(z), \quad \forall i \in [m].
\end{aligned} \tag{E.7}$$

If $y_\ell = \arg\max_{y \in \mathcal{Z} \,:\, \epsilon_\ell \leq \min_i \Delta^i_{\text{safe}}(y)} y^\top \theta_*$ is in $\mathcal{Y}_\ell$, then we are done. Assume then that $y_\ell \notin \mathcal{Y}_\ell$. By construction, since $\Delta^i_{\text{safe}}(y_\ell) > 0$ for all i, $\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) = 0$. Using that (E.7) is a sufficient condition for inclusion in $\mathcal{Y}_\ell$, this implies that

$$\exists i \in [m] \quad \text{s.t.} \quad (3c_d + 3c_e + 6c_dc_3 + 12c_dc_4 + c_4) \left(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^\ell(y_\ell)) + \min_j |\Delta^j_{\text{safe}}(y_\ell)| \right) > \Delta^i_{\text{safe}}(y_\ell).$$

which implies

$$\exists i \in [m] \quad \text{s.t.} \quad (3c_d + 3c_e + 6c_d c_3 + 12 c_d c_4 + c_4) \left(\epsilon_\ell + \mathfrak{p}(\widetilde{\Delta}^\ell(y_\ell)) + |\Delta^i_{\text{safe}}(y_\ell)| \right) > \Delta^i_{\text{safe}}(y_\ell). \quad \text{(E.8)}$$

By construction, though, $\Delta^i_{\text{safe}}(y_\ell) \geq \epsilon_\ell$. If we assume that

$$3c_d + 3c_e + 6c_d c_3 + 12 c_d c_4 + c_4 \leq 1/4, \quad \text{(E.9)}$$

then (E.8) can only hold if $\mathfrak{p}(\widetilde{\Delta}^\ell(y_\ell)) > 0$. This implies that $\max_{y \in \mathcal{Y}_\ell} y^\top \theta_* > y_\ell^\top \theta_*$. Thus, in this case,

$$\widetilde{\Delta}^\ell(z) = \max_{y \in \mathcal{Y}_\ell} y^\top \theta_* - z^\top \theta_* > y_\ell^\top \theta_* - z^\top \theta_* = \Delta^{\epsilon_\ell}(z)$$

which proves the result. □

Lemma 70. *On $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$, for all $z \in \mathcal{Y}_{\text{end}}$ we have*

$$\Delta^i_{\text{safe}}(z) \geq -c_g \epsilon, \quad \forall i \in [m],$$
$$\widehat{\Delta}^{i,\iota_\epsilon}_{\text{safe}}(z) \geq (3c_d + 3c_e - c_g)\epsilon, \quad \forall i \in [m].$$

Furthermore, $z_* \in \mathcal{Y}_{\text{end}}$.

Proof. Recall that

$$\mathcal{Y}_{\text{end}} = \{ z \in \mathcal{Z} \ : 3c_d \left(\min_j |\widehat{\Delta}^{j,\iota_\epsilon}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\iota_\epsilon}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\iota_\epsilon}(z)) \right)$$
$$+ 3(c_d + c_e)\epsilon - c_g \epsilon \leq \widehat{\Delta}^{i,\iota_\epsilon}_{\text{safe}}(z), \forall i \in [m] \}$$

On $\mathcal{E}_{\text{safe}}$, we have

$$\widehat{\Delta}^{i,\iota_\epsilon}_{\text{safe}}(z) \leq \Delta^i_{\text{safe}}(z) + 3c_d \left(\min_j |\widehat{\Delta}^{j,\iota_\epsilon}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\iota_\epsilon}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\iota_\epsilon}(z)) \right) + 3(c_d + c_e)\epsilon$$

so it follows that if $z \in \mathcal{Y}_{\text{end}}$, then

$$-c_g \epsilon \leq \Delta^i_{\text{safe}}(x).$$

To see that $z_* \in \mathcal{Y}_{\text{end}}$, note that by definition of $\mathcal{Y}_{\text{end}}$, using a calculation analogous to (E.7), a sufficient condition for $z \in \mathcal{Y}_{\text{end}}$ is

$$(3c_d + 3c_e + 6c_d c_3 + 12 c_d c_4 + c_4 - c_g)\epsilon + (3c_d + 3c_d c_3 + 6 c_d c_4 + c_4)\mathfrak{p}(\widetilde{\Delta}^{\iota_\epsilon}(z))$$
$$+ (3c_d + 6c_d c_4 + c_4) \min_j |\Delta^j_{\text{safe}}(z)| + (3c_d + 3c_d c_3 + 6c_d c_4 + c_4) \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z))$$
$$\leq \Delta^i_{\text{safe}}(z), \quad \forall i \in [m].$$

By definition of z_* and since, by Lemma 67, all $z \in \mathcal{Y}_{\iota_\epsilon}$ are safe, we have $\Delta^{\epsilon_{\iota_\epsilon}}(z_*) \leq 0$. Furthermore, by definition we also have $\Delta^j_{\text{safe}}(z_*) \geq 0$ for all j, so $\mathfrak{p}(-\Delta^j_{\text{safe}}(z_*)) = 0$. Thus, assuming that

$$3c_d + 3c_e + 6c_d c_3 + 12 c_d c_4 + c_4 - c_g \leq 0 \quad \text{(E.10)}$$

a sufficient condition to guarantee $z_* \in \mathcal{Y}_{\text{end}}$ is that

$$(8c_d + 16 c_d c_4 + c_4) \min_j |\Delta^j_{\text{safe}}(z_*)| \leq \Delta^i_{\text{safe}}(z_*), \quad \forall i \in [m].$$

However, as long as

$$3c_d + 6c_d c_4 + c_4 \leq 1, \quad \text{(E.11)}$$

this is true, since by definition $\Delta^i_{\text{safe}}(z_*) \geq 0$. □

E.4.5 Algorithm Correctness and Sample Complexity

Lemma 71 (Correctness)**.** *On $\mathcal{E}_{\text{safe}} \cap \mathcal{E}_{\text{RAGE}^\epsilon}$, we will have that*

$$\widehat{z}^\top \theta_* \geq (z_*)^\top \theta_* - \frac{c_3(1+c_g)}{1-c_3}\epsilon, \quad \Delta^i_{\text{safe}}(\widehat{z}) \geq -c_g\epsilon, \forall i \in [m].$$

Proof. We choose $\widehat{z}$ to be any $z \in \mathcal{Y}_{\text{end}}$ such that $\widehat{\Delta}^{\text{end}}(z) = 0$. By Lemma 70, we have that $\Delta^i_{\text{safe}}(\widehat{z}) \geq -c_g\epsilon$ for all $i \in [m]$. If $\widetilde{\Delta}^{\text{end}}(\widehat{z}) \leq 0$, we are done, since by Lemma 70, $z_* \in \mathcal{Y}_{\text{end}}$, so $\widehat{z}^\top\theta_* \geq (z_*)^\top\theta_*$. Assume that $\widetilde{\Delta}^{\text{end}}(\widehat{z}) > 0$. By Lemma 68, we have that

$$\widetilde{\Delta}^{\text{end}}(\widehat{z}) \leq c_3\epsilon + c_3\mathfrak{p}(\widetilde{\Delta}^{\text{end}}(\widehat{z})) + c_3 \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(\widehat{z})).$$

By Lemma 70, since $\widehat{z} \in \mathcal{Y}_{\text{end}}$, $\mathfrak{p}(-\Delta^j_{\text{safe}}(\widehat{z})) \leq c_g\epsilon$ for all j, so we can bound

$$\widetilde{\Delta}^{\text{end}}(\widehat{z}) \leq c_3(1+c_g)\epsilon + c_3\mathfrak{p}(\widetilde{\Delta}^{\text{end}}(\widehat{z})) = c_3(1+c_g)\epsilon + c_3\widetilde{\Delta}^{\text{end}}(\widehat{z}).$$

We can rearrange this as

$$\widetilde{\Delta}^{\text{end}}(\widehat{z}) \leq \frac{c_3(1+c_g)}{1-c_3}\epsilon$$

which proves the result, since, by Lemma 70, $\Delta^{\text{end}}(\widehat{z}) = \max_{y\in\mathcal{Y}_{\text{end}}} y^\top\theta_* - \widehat{z}^\top\theta_* \geq (z_*)^\top\theta_* - \widehat{z}^\top\theta_*$. □

Lemma 72. *On $\mathcal{E}_{\text{RAGE}^\epsilon} \cap \mathcal{E}_{\text{safe}}$, the total complexity of Line 6 is bounded by*

$$C \cdot \sum_{\ell=1}^{\iota_\epsilon} \inf_{\lambda\in\triangle_\mathcal{X}} \max_{z\in\mathcal{Z}} \frac{\|z\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\left(\min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + \mathfrak{p}(\Delta^{\epsilon_{\ell-1}}(z)) + \epsilon_\ell\right)^2} + 4\iota_\epsilon \log(\tfrac{4m|\mathcal{Z}|\iota_\epsilon^2}{\delta})$$

for an absolute constant C.

Proof. Applying the same argument as in Claim 1 but in the opposite direction, we have

$$\begin{aligned}
&\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) \\
&\geq -2(c_3+2c_4)\epsilon_\ell + (1-c_3-2c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + (1-2c_4)\min_j |\Delta^j_{\text{safe}}(z)| + (1-c_3-2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)).
\end{aligned}$$

We assume that c_3, c_4, and c_0 are chosen such that

$$1 - 2c_3 - 4c_4 \geq c_0, \tag{E.12}$$

which allows us to bound:

$$\begin{aligned}
&\inf_{\lambda\in\triangle_\mathcal{X}} \max_{z\in\mathcal{Z}} -c_d\left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell\right) + \sqrt{\frac{\|z\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau}} \\
&\leq \inf_{\lambda\in\triangle_\mathcal{X}} \max_{z\in\mathcal{Z}} -c_dc_0\left(\min_j |\Delta^j_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) + \epsilon_\ell\right) + \sqrt{\frac{\|z\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau}}.
\end{aligned}$$

It follows that if, for all $z \in \mathcal{Z}$,

$$\tau \geq \frac{\|z\|^2_{A(\lambda)^{-1}}}{\left(c_d c_0 \min_j |\Delta^j_{\text{safe}}(z)| + c_d c_0 \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + c_d c_0 \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) + (c_d c_0 + c_e)\epsilon_\ell\right)^2} \cdot \log \tfrac{4m|\mathcal{Z}|\ell^2}{\delta}$$

we will have that this is less than $c_e \epsilon_\ell$. Since we can take the best-case $\lambda \in \triangle_{\mathcal{X}}$, and since τ_ℓ is always within a factor of 2 of the optimal, it follows that

$$\tau_\ell \leq \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} \frac{2\|z\|^2_{A(\lambda)^{-1}} \cdot \log \frac{4m|\mathcal{Z}|\ell^2}{\delta}}{\left(c_d c_0 \min_j |\Delta^j_{\text{safe}}(z)| + c_d c_0 \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + c_d c_0 \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) + (c_d c_0 + c_e)\epsilon_\ell\right)^2} + 4 \log \tfrac{4m|\mathcal{Z}|\ell^2}{\delta}$$

The result then follows by summing over epochs and lower bounding $\widetilde{\Delta}^{\ell-1}(z)$ by $\Delta^{\epsilon_{\ell-1}}(z)$ using Lemma 69, and assuming that

$$c_d c_0 + c_e \geq c_0. \tag{E.13}$$

□

Proof of Theorem 31. By Lemma 65 we have that $\mathcal{E}_{\text{safe}}$ holds with probability at least $1 - \delta$. By Lemma 60, we have that $\mathcal{E}^\ell_{\text{RAGE}^\epsilon}$ holds with probability at least $1 - \delta/(4\ell^2)$. It follows then that $\mathcal{E}_{\text{safe}} \cup (\cup_\ell \mathcal{E}^\ell_{\text{RAGE}^\epsilon})$ holds with probability at least

$$1 - \delta - \sum_\ell \frac{\delta}{4\ell^2} \geq 1 - 2\delta.$$

Assume henceforth that $\mathcal{E}_{\text{safe}} \cup (\cup_\ell \mathcal{E}^\ell_{\text{RAGE}^\epsilon})$ holds. Equation (E.5) follows by Lemma 71. The total number of samples collected on Line 6 can be bounded by Lemma 72. It remains to bound the total number of samples used by RAGE$^\epsilon$.

By Lemma 64, at epoch ℓ RAGE$^\epsilon$ will collect at most

$$C\lceil \log \frac{2}{c_f \epsilon_\ell} \rceil \cdot \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} \frac{\|z - y^\ell_\star\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{8|\mathcal{Z}|^2 \log^4(1/\epsilon)}{\delta})}{(\max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \epsilon_\ell)^2} + 8\lceil \log \tfrac{2}{c_f \epsilon} \rceil \log(\frac{4|\mathcal{Z}|^2 \lceil \log \frac{2}{c_f \epsilon} \rceil^2}{\delta})$$

samples, where $y^\ell_\star = \arg\max_{y \in \mathcal{Y}_\ell} y^\top \theta_*$. Assume that $\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) > 0$, then we can upper bound $\min_j |\Delta^j_{\text{safe}}(z)| \leq \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z))$, and by Lemma 68 we can lower bound

$$\max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) \geq (1 - 2c_4) \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) - c_4 \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) - c_4 \epsilon_{\ell-1}.$$

Assume instead that $\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) = 0$. Then again by Lemma 68:

$$\max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) \geq 0 = \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) \geq (1 - 2c_4) \max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) - c_4 \mathfrak{p}(\widetilde{\Delta}^{\ell-1}(z)) - c_4 \epsilon_{\ell-1}.$$

By Lemma 67, it follows that

$$\max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \epsilon_\ell$$

$$\geq (1-2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + (1-c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + (1-2c_4)\epsilon_\ell.$$

By definition and Lemma 67 and Lemma 70 for all ℓ including $\ell = \text{end}$, we can bound $\mathfrak{p}(-\Delta^j_{\text{safe}}(y^\ell_\star)) \leq c_g\epsilon$. Furthermore, by definition $\mathfrak{p}(\widetilde{\Delta}^\ell(y^\ell_\star)) = 0$. Putting all of this together, we have:

$$\begin{aligned}
&\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}} \frac{\|z-y^\ell_\star\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{8|\mathcal{Z}|^2\log^4(1/\epsilon)}{\delta})}{(\max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widetilde{\Delta}^\ell(z)) + \epsilon_\ell)^2} \\
&\leq \inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}} \frac{\|z-y^\ell_\star\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{8|\mathcal{Z}|^2\log^4(1/\epsilon)}{\delta})}{((1-2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + (1-c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + (1-2c_4)\epsilon_\ell)^2} \\
&\leq \inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}} \frac{2\|z-z_*\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{8|\mathcal{Z}|^2\log^4(1/\epsilon)}{\delta})}{((1-2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + (1-c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + (1-2c_4)\epsilon_\ell)^2} \\
&\quad + \inf_{\lambda\in\triangle_{\mathcal{X}}} \frac{2\|z_*-y^\ell_\star\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{8|\mathcal{Z}|^2\log^4(1/\epsilon)}{\delta})}{((1-2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(y^\ell_\star)) + (1-c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(y^\ell_\star)) + (1-2c_4-c_g)\epsilon_\ell)^2} \\
&\leq \inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}} \frac{4\|z-z_*\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{8|\mathcal{Z}|^2\log^4(1/\epsilon)}{\delta})}{((1-2c_4)\max_j \mathfrak{p}(-\Delta^j_{\text{safe}}(z)) + (1-c_4)\mathfrak{p}(\widetilde{\Delta}^\ell(z)) + (1-2c_4-c_g)\epsilon_\ell)^2}
\end{aligned}$$

As long as

$$1-2c_4-c_g \geq c_0, \tag{E.14}$$

summing over the epochs and lower bounding $\widetilde{\Delta}^\ell(z)$ by $\Delta^{\epsilon_\ell}(z)$ via Lemma 69 gives the result. Finally, the settings of the constants follows from Lemma 59. □

E.4.6 Proofs of Corollaries to Theorem 16

Proof of Corollary 2. If $m = 1$, $\mu_{*,1} = 0$, and $\gamma = 1$, then we have $\Delta_{\text{safe}}(z) = 1$ for each z, and $\Delta^{\widetilde{\epsilon}}(z) = \Delta(z)$ for $\epsilon \leq 1$. The result follows directly from this and some algebra. □

Proof of Corollary 3. We can trivially upper bound the complexity given in Theorem 16 by

$$\begin{aligned}
&C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{m|\mathcal{Z}|}{\delta})}{\epsilon^2} + C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z-z_*\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{|\mathcal{Z}|}{\delta})}{\epsilon^2} + C_0 \\
&\leq C\cdot\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\frac{\|z\|^2_{A(\lambda)^{-1}}\cdot\log(\frac{m|\mathcal{Z}|}{\delta})}{\epsilon^2} + C_0.
\end{aligned}$$

In the case when $\mathcal{X} = \mathcal{Z}$, we can bound $\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\mathcal{Z}}\|z\|^2_{A(\lambda)^{-1}} \leq d$ by Kiefer-Wolfowitz (Lattimore and Szepesvári, 2020), which proves the result. □

E.5 Computationally Efficient Optimization

Throughout, we will let $\mathcal{R}(z;\xi_1,\dots,\xi_n)$ denote some generic weighted risk estimate of the form

$$\mathcal{R}(z;\xi_1,\dots,\xi_n) = \sum_{t=1}^T f_t(\xi_1,\dots,\xi_n)\mathcal{I}\{z(u_t)\neq v_t\}$$

for some weights $f_t(\xi_1, \dots, \xi_n)$ and observations (u_t, v_t). The exact setting of $\mathcal{R}$ will change from line to line—we simply use it as a stand-in for an objective that a cost-sensitive-classification oracle can efficiently minimize. We will also use $f(\xi_1, \dots, \xi_n)$ to refer to some generic function (the particular form of which is not important).

Lemma 73. *Consider some $z, \widetilde{z} \in \triangle_{\mathcal{H}}$. Denote*

$$\rho_\lambda(h, h') = \mathbb{E}_{U\sim\nu}\left[\frac{\mathcal{I}\{h(U) \neq h'(U)\}}{\lambda(U)/\nu(U)}\right] = \|h - h'\|^2_{A(\lambda)^{-1}}$$

and overload notation so that $z = \sum_{h\in\mathcal{H}} z_h h$ denotes the feature vector for the mixed classifier z. Then,

$$\sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} \rho_\lambda(h, h') = \mathbb{E}_{U\sim\nu}\left[\frac{(z(U) - \widetilde{z}(U))^2}{\lambda(U)/\nu(U)}\right] = \|z - \widetilde{z}\|^2_{A(\lambda)^{-1}}.$$

Proof. Note that

$$\rho_\lambda(h, h') = \mathbb{E}_{U\sim\nu}\left[\frac{\mathcal{I}\{h(U) \neq h'(U)\}}{\lambda(U)/\nu(U)}\right] = \mathbb{E}_{U\sim\nu}\left[\frac{|h(U) - h'(U)|}{\lambda(U)/\nu(U)}\right] = \mathbb{E}_{U\sim\nu}\left[\frac{(h(U) - h'(U))^2}{\lambda(U)/\nu(U)}\right]$$

where the final equality holds because $|h(U) - h'(U)|$ is always either 0 or 1. Thus,

$$\begin{aligned}
\sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} \rho_\lambda(h, h') &= \sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} \mathbb{E}_{U\sim\nu}\left[\frac{(h(U) - h'(U))^2}{\lambda(U)/\nu(U)}\right] \\
&= \mathbb{E}_{U\sim\nu}\left[\frac{\sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} (h(U) - h'(U))^2}{\lambda(U)/\nu(U)}\right] \\
&= \mathbb{E}_{U\sim\nu}\left[\frac{\sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} (h(U) + h'(U) - 2h(U)h'(U))}{\lambda(U)/\nu(U)}\right]. \qquad \text{(E.15)}
\end{aligned}$$

However,

$$\sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} h(U) = \sum_{h\in\mathcal{H}} z_h h(U) = z(U), \quad \sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} h'(U) = \widetilde{z}(U)$$

and

$$\sum_{h,h'\in\mathcal{H}} z_h \widetilde{z}_{h'} h(U) h'(U) = (\sum_{h\in\mathcal{H}} z_h h(U))(\sum_{h'\in\mathcal{H}} \widetilde{z}_{h'} h'(U)) = z(U)\widetilde{z}(U).$$

Thus,

$$\text{(E.15)} = \mathbb{E}_{U\sim\nu}\left[\frac{(z(U) - \widetilde{z}(U))^2}{\lambda(U)/\nu(U)}\right]$$

which proves the first equality. To prove the second, recall that $[h]_u = \nu(u)h(u)$, so $[z]_u = \sum_{h\in\mathcal{H}} z_h [h]_u = \nu(u)z(u)$. It follows that,

$$\|z - \widetilde{z}\|^2_{A(\lambda)^{-1}} = \sum_u \frac{\nu(u)^2}{\lambda(u)}(z(u) - \widetilde{z}(u))^2 = \mathbb{E}_{U\sim\nu}\left[\frac{(z(U) - \widetilde{z}(U))^2}{\lambda(U)/\nu(U)}\right]$$

which proves the second equality. □

E.5.1 Computational Efficiency of RAGE$^\epsilon$

RAGE$^\epsilon$ requires solving the optimization

$$\inf_{\lambda\in\triangle_{\mathcal{X}}}\max_{z\in\triangle_{\mathcal{H}}}\min_{\alpha\in\mathcal{A}} -c_a(\mathfrak{p}(-\widehat{\Delta}_{\text{safe}}(z))+\mathfrak{p}(\widehat{\Delta}^{\ell-1}(z))+\epsilon_\ell)+\alpha\|z-\widehat{y}_{\ell-1}\|^2_{A(\lambda)^{-1}}+\frac{\log(2|\mathcal{Z}|^2|\mathcal{A}|\ell^2/\delta)}{\alpha\tau}. \tag{E.16}$$

Here we take τ to be fixed, and recall that

$$\widehat{y}_\ell \leftarrow \text{argmin}_{y\in\mathcal{Y}}\min_{\alpha\in\mathcal{A}}\widetilde{R}^\alpha_\ell(y)-\widetilde{R}^\alpha_\ell(\widehat{y}_{\ell-1})+2\alpha\|y-\widehat{y}_{\ell-1}\|^2_{A(\lambda_\ell)^{-1}}+\frac{2\log(2|\mathcal{Z}|^2|\mathcal{A}|\ell^2/\delta)}{\alpha\tau_\ell}$$
$$\widehat{\Delta}^\ell(y)\leftarrow\min_{\alpha\in\mathcal{A}}\widetilde{R}^\alpha_\ell(y)-\widetilde{R}^\alpha_\ell(\widehat{y}_\ell)+\alpha\|y-\widehat{y}_\ell\|^2_{A(\lambda_\ell)^{-1}}+\frac{\log(2|\mathcal{Z}|^2|\mathcal{A}|\ell^2/\delta)}{\alpha\tau_\ell}.$$

Furthermore, $\mathcal{Y}$ will be a set of the form

$$\bigcup_{k=1}^{\ell'}\mathcal{Y}_k=\bigcup_{k=1}^{\ell'}\left\{z\in\mathcal{Z}\ :\ c(\epsilon_k+\mathfrak{p}(\widehat{\Delta}^{k-1}(z))+\max_{j\in[n]}\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(z))+\min_{j\in[n]}|\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|)\le\widehat{\Delta}^{i,k}_{\text{safe}}(h),\forall i\in[n]\right\}$$

Recall also that

$$\|h-h'\|^2_{A(\lambda)^{-1}}=\mathbb{E}_{U\sim\nu}\left[\frac{\mathcal{I}\{h(U)\neq h'(U)\}}{(9\lambda(U)/10+1/10d)/\nu(U)}\right]=\sum_{U\in\mathcal{X}}\frac{\nu(U)^2}{9\lambda(U)/10+1/10d}\mathcal{I}\{h(U)\neq h'(U)\}$$

and

$$\widetilde{R}^\alpha_\ell(h)=\frac{1}{\tau_\ell}\sum_{t=1}^{\tau_\ell}\frac{1}{w_t+\alpha}\mathcal{I}\{h(u_t)\neq v_t\}.$$

For $z\in\triangle_{\mathcal{H}}$, we denote $\widetilde{R}^\alpha_\ell(z)=\sum_{h\in\mathcal{H}}z_h\widetilde{R}^\alpha_\ell(h)$ and $\mathcal{R}(z;\alpha)=\sum_{h\in\mathcal{H}}z_h\mathcal{R}(h;\alpha)$. Finally, we assume that $\widehat{\Delta}_{\text{safe}}(z)=\min_{\alpha\in\mathcal{A}}\mathcal{R}(z;\alpha)+f(\alpha)$.

Solving for $\widehat{y}_\ell$

Using Lemma 73, we can write the optimization for $\widehat{y}_\ell$ as

$$\min_{k\in[\ell']}\min_{y\in\mathcal{Y}_k}\min_{\alpha\in\mathcal{A}}\frac{1}{\tau_\ell}\sum_{t=1}^{\tau_\ell}\frac{1}{w_t+\alpha}\sum_{h\in\mathcal{H}}y_h\mathcal{I}\{h(u_t)\neq v_t\}+\alpha\sum_{h,h'\in\mathcal{H}}y_h\widehat{y}_{\ell-1,h'}\sum_{U\in\mathcal{X}}\frac{\nu(U)^2}{9\lambda_\ell(U)/10+1/10d}\mathcal{I}\{h(U)\neq h'(U)\}$$
$$-\widetilde{R}^\alpha_\ell(\widehat{y}_{\ell-1})+\frac{\log(2|\mathcal{Z}|^2|\mathcal{A}|\ell^2/\delta)}{\alpha\tau_\ell}$$

We can rewrite

$$\sum_{h,h'\in\mathcal{H}}y_h\widehat{y}_{\ell-1,h'}\sum_{U\in\mathcal{X}}\frac{\nu(U)^2}{9\lambda_\ell(U)/10+1/10d}\mathcal{I}\{h(U)\neq h'(U)\}=\sum_{h\in\mathcal{H}}y_h\sum_{i=1}^{\|\widehat{y}_{\ell-1}\|_0|\mathcal{X}|}w_i\mathcal{I}\{h(u_i)\neq v_i\}$$

for some weights w_i. It follows that if $\|\widehat{y}_{\ell-1}\|_0$ is polynomial in problem parameters then the optimization for $\widehat{y}_\ell$ can be written as

$$\min_{k\in[\ell']}\min_{y\in\mathcal{Y}_k}\min_{\alpha\in\mathcal{A}}\mathcal{R}(y;\alpha)+f(\alpha)$$

for $\mathcal{R}(y;\alpha)$ a CSC loss over only polynomially many points (as well as linear in y and convex in α), and $f(\alpha)$ convex in α. Note also that, for any y, we can upper bound $\mathcal{R}(y;\alpha)\leq\mathcal{O}(\frac{1}{\alpha}+d\alpha)$. Here $\mathcal{Y}_k$ a set of the form

$$\left\{z\in\triangle_{\mathcal{H}}\ :\ \sum_{h\in\mathcal{H}}z_h c\Big(\epsilon_k+\mathfrak{p}(\widehat{\Delta}^{k-1}(h))+\max_{j\in[n]}\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(h))+\min_{j\in[n]}|\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|\Big)\leq\sum_{h\in\mathcal{H}}z_h\widehat{\Delta}^{i,k}_{\text{safe}}(h),\forall i\in[n]\right\}$$

$\widehat{y}_\ell$ will be the element in $\mathcal{Y}_k$ minimizing the, for the k achieving the minimum. The dual of this problem has the form

$$\begin{aligned}&\min_{k\in[\ell']}\min_{z\in\triangle_{\mathcal{H}}}\min_{\alpha\in\mathcal{A}}\max_{\mu_i\geq0,i\in[n]}\mathcal{R}(z;\alpha)+f(\alpha)\\&\quad+\sum_{i=1}^n\mu_i\left(\sum_{h\in\mathcal{H}}z_h c\Big(\epsilon_k+\mathfrak{p}(\widehat{\Delta}^{k-1}(h))+\max_{j\in[n]}\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(h))+\min_{j\in[n]}|\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|\Big)-\sum_{h\in\mathcal{H}}z_h\widehat{\Delta}^{i,k}_{\text{safe}}(h)\right).\end{aligned}$$

Note that we can swap the min over α and z without issue. Furthermore, for a fixed μ, the objective is linear in z, and for a fixed z, the objective is linear in μ. By the minimax theorem, we can then swap the min and max to obtain the equivalent optimization:

$$\begin{aligned}&\min_{k\in[\ell']}\min_{\alpha\in\mathcal{A}}\max_{\mu_i\geq0,i\in[n]}\min_{z\in\triangle_{\mathcal{H}}}\mathcal{R}(z;\alpha)+f(\alpha)\\&\quad+\sum_{i=1}^n\mu_i\left(\sum_{h\in\mathcal{H}}z_h c\Big(\epsilon_k+\mathfrak{p}(\widehat{\Delta}^{k-1}(h))+\max_{j\in[n]}\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(h))+\min_{j\in[n]}|\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|\Big)-\sum_{h\in\mathcal{H}}z_h\widehat{\Delta}^{i,k}_{\text{safe}}(h)\right).\end{aligned}$$

We can simply enumerate over k and α, as there are a finite number of each of these constraints. For a fixed k and α, to solve the inner maxmin problem, we can apply the approach proposed in (Agarwal et al., 2018). In particular, we alternate between running the exponential gradient algorithm for the μ player, and computing the best-response for the z player. The update to the μ player is trivial, as the problem is simply linear in μ (in practice, as in (Agarwal et al., 2018), we will also upper bound the domain of μ_i by some value B, to ensure this is finite).

Computing the best-response for the z player (with μ fixed) is slightly trickier. Ignoring all other parameters, which are all currently fixed, the minimization over z can be written as

$$\begin{aligned}&\min_{z\in\triangle_{\mathcal{H}}}\sum_{h\in\mathcal{H}}z_h\sum_t a_t\mathcal{I}\{h(u_t)\neq o_t\}\\&+\sum_{i=1}^n\mu_i\left(\sum_{h\in\mathcal{H}}z_h c\Big(\epsilon_k+\mathfrak{p}(\widehat{\Delta}^{k-1}(h))+\max_{j\in[n]}\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(h))+\min_{j\in[n]}|\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|\Big)-\sum_{h\in\mathcal{H}}z_h\widehat{\Delta}^{i,k}_{\text{safe}}(h)\right).\\&=\min_{z\in\triangle_{\mathcal{H}}}\sum_{h\in\mathcal{H}}z_h\Bigg(\sum_t a_t\mathcal{I}\{h(u_t)\neq o_t\}\\&\qquad+\sum_{i\in[n]}c_i\Big(\epsilon_k+\mathfrak{p}(\widehat{\Delta}^{k-1}(h))+\max_{j\in[n]}\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(h))+\min_{j\in[n]}|\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|-\widehat{\Delta}^{i,k}_{\text{safe}}(h)\Big)\Bigg).\end{aligned}$$

Now note that $\max_{j\in[n]} \mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(z)) = \sup_{\widetilde{\lambda}\in\triangle_n} \sum_{j\in[n]} \widetilde{\lambda}_j \mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(z))$, and similarly for $\min_{j\in[n]} |\widehat{\Delta}^{j,k-1}_{\text{safe}}(z)|$. Using this, we can rewrite the above optimization as

$$\min_{z\in\triangle_{\mathcal{H}}} \max_{\widetilde{\lambda}^{1h}\in\triangle_n, h\in\mathcal{H}} \min_{\widetilde{\lambda}^{2h}\in\triangle_n, h\in\mathcal{H}} \sum_{h\in\mathcal{H}} z_h \Bigg(\sum_t a_t \mathcal{I}\{h(u_t)\neq o_t\} \\ + \sum_{i\in[n]} c_i \Big(\epsilon_k + \mathfrak{p}(\widehat{\Delta}^{k-1}(h)) + \sum_{j\in[n]} \widetilde{\lambda}^{1h}_j \mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)) + \sum_{j\in[n]} \widetilde{\lambda}^{2h}_j |\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)| - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big) \Bigg).$$

We also have:

$$\mathfrak{p}(-\widehat{\Delta}^{j,k-1}_{\text{safe}}(z)) = \max_{\beta\in[0,1]} -\beta\widehat{\Delta}^{j,k-1}_{\text{safe}}(z), \quad |\widehat{\Delta}^{j,k-1}_{\text{safe}}(z)| = \max_{\beta\in[-1,1]} \beta\widehat{\Delta}^{j,k-1}_{\text{safe}}(z).$$

So we can further simplify the above to:

$$\min_{z\in\triangle_{\mathcal{H}}} \max_{\widetilde{\lambda}^{1h}\in\triangle_n, h\in\mathcal{H}} \min_{\widetilde{\lambda}^{2h}\in\triangle_n, h\in\mathcal{H}} \max_{\beta_1^h, \beta_2^{hj}\in[0,1], \beta_3^{hj}\in[-1,1], h\in\mathcal{H}} \sum_{h\in\mathcal{H}} z_h \Bigg(\sum_t a_t \mathcal{I}\{h(u_t)\neq o_t\} \\ + \sum_{i\in[n]} c_i \Big(\epsilon_k + \beta_1^h \widehat{\Delta}^{k-1}(h) - \sum_{j\in[n]} \widetilde{\lambda}^{1h}_j \beta_2^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) + \sum_{j\in[n]} \widetilde{\lambda}^{2h}_j \beta_3^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big) \Bigg).$$

Note that the objective is linear in β and $\widetilde{\lambda}^2$, and both have continuous, compact, convex constraint sets, so we can swap the min and max to get that the above is equivalent to

$$\min_{z\in\triangle_{\mathcal{H}}} \max_{\widetilde{\lambda}^{1h}\in\triangle_n, h\in\mathcal{H}} \max_{\beta_1^h, \beta_2^{hj}\in[0,1], \beta_3^{hj}\in[-1,1], h\in\mathcal{H}} \min_{\widetilde{\lambda}^{2h}\in\triangle_n, h\in\mathcal{H}} \sum_{h\in\mathcal{H}} z_h \Bigg(\sum_t a_t \mathcal{I}\{h(u_t)\neq o_t\} \\ + \sum_{i\in[n]} c_i \Big(\epsilon_k + \beta_1^h \widehat{\Delta}^{k-1}(h) - \sum_{j\in[n]} \widetilde{\lambda}^{1h}_j \beta_2^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) + \sum_{j\in[n]} \widetilde{\lambda}^{2h}_j \beta_3^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big) \Bigg).$$

We can write this in the form

$$\min_{z\in\triangle_{\mathcal{H}}} \max_{\widetilde{\lambda}^1,\beta} g(z;\widetilde{\lambda}^1,\beta) \tag{E.17}$$

for

$$g(z;\widetilde{\lambda}^1,\beta) := \min_{\widetilde{\lambda}^{2h}\in\triangle_n, h\in\mathcal{H}} \sum_{h\in\mathcal{H}} z_h \Bigg(\sum_t a_t \mathcal{I}\{h(u_t)\neq o_t\} \\ + \sum_{i\in[n]} c_i \Big(\epsilon_k + \beta_1^h \widehat{\Delta}^{k-1}(h) - \sum_{j\in[n]} \widetilde{\lambda}^{1h}_j \beta_2^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) + \sum_{j\in[n]} \widetilde{\lambda}^{2h}_j \beta_3^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big) \Bigg).$$

To solve this, we will apply a version of Frank-Wolfe that handles adversarial losses to the outer player (see Section 4.2 of (Hazan and Kale, 2012)), and will play best response for the inner player.

From the perspective of the outer player, at iteration t of the algorithm given in (Hazan and Kale, 2012), they must optimize the function

$$f_t(z) = g(z;\widetilde{\lambda}^1_t,\beta_t) = \sum_{h\in\mathcal{H}} z_h c_h(\widetilde{\lambda}^1_t,\beta_t)$$

for some $c_h(\widetilde{\lambda}^1_t, \beta_t)$. Note that this is $L = \max_h |c_h(\widetilde{\lambda}^1_t, \beta_t)|$ Lipschitz in the ℓ_1-norm, and that we can bound this L for all t by something like $\mathcal{O}(\frac{1}{\alpha} + d\alpha + n)$. The algorithm introduced in Section 4.2 of (Hazan and Kale, 2012) computes the standard FW update

$$\widetilde{z}_t = \text{argmin}_{z \in \triangle_{\mathcal{H}}} \nabla F_t(z_t)^\top z, \quad z_{t+1} = (1 - t^{-1/4}) z_t + t^{-1/4} \widetilde{z}_t$$

for

$$F_t(z) = \frac{1}{t} \sum_{\tau=1}^{t} \nabla f_\tau(z_\tau)^\top z + \sigma_t \|z - z_1\|_2^2$$

for $\sigma_t = (L/D) t^{-1/4}$ for $D = \max_{z_1, z_2 \in \triangle_{\mathcal{H}}} \|z_1 - z_2\|_1$ (note that in that work, the function seems to be Lipschitz in the ℓ_2 norm while here we use ℓ_1—this does not seem to change their result at all). It is shown in (Hazan and Kale, 2012) that running this procedure we obtain the bound, for any $z \in \triangle_{\mathcal{H}}$,

$$\sum_{t=1}^{T} (f_t(z_t) - f_t(z)) \leq 57 L D T^{3/4}.$$

It follows that if we are able to compute $\widetilde{z}_t$ efficiently, and if the max player plays best response (and the best response can be computed efficiently), using analysis similar to that in (Agarwal et al., 2018), we can show that an approximate solution to (E.17) will be found in a polynomial number of iterations.

Computing the Best Response for $\widetilde{\lambda}^1, \beta$. For the inner player, they must solve

$$\max_{\widetilde{\lambda}^1, \beta} g(z_t; \widetilde{\lambda}^1, \beta).$$

Assume that $\|\widetilde{z}_t\|_0 \leq m$ for each t, and that $\|z_1\|_0 = 1$. Then z_t will be $(mt+1)$-sparse, so the sum in $g(z_t; \widetilde{\lambda}^1, \beta)$ will contain at most $(mt+1)$ values. Note that the optimization over β^h and $\widetilde{\lambda}^{1h}$ is completely independent, so to compute the best-response, we need to solve the following problem at most $(mt+1)$ times:

$$\max_{\widetilde{\lambda}^{1h} \in \triangle_n} \max_{\beta_1^h, \beta_2^{hj} \in [0,1], \beta_3^{hj} \in [-1,1]} \min_{\widetilde{\lambda}^{2h} \in \triangle_n} \sum_{i \in [n]} c_i \Big(\beta_1^h \widehat{\Delta}^{k-1}(h) - \sum_{j \in [n]} \widetilde{\lambda}_j^{1h} \beta_2^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) + \sum_{j \in [n]} \widetilde{\lambda}_j^{2h} \beta_3^{hj} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) \Big).$$

The optimization over the first two terms is trivial and can be solved by enumerating. The third term now is a maxmin problem, however, this can also be solved trivially as it is equivalent to $\min_{j \in [n]} |\widehat{\Delta}^{j,k-1}_{\text{safe}}(h)|$. Note that each of these gap terms is themself the solution to an optimization over $\alpha \in \mathcal{A}$, but that can be solved easily for each (since there are at most polynomial of them), so they can be regarded as constants.

Thus, we conclude that the best response for $\widetilde{\lambda}^1, \beta$ can be computed efficiently, assuming that m is polynomial in problem parameters. Note that the values of β^h and $\widetilde{\lambda}^{1h}$ do not matter for $h \notin \text{support}(z_t)$ do not matter to compute the best response, so we can set them to the same value for all $h \notin \text{support}(z_t)$.

Computing $\widetilde{z}_t$. It remains to show that we can efficiently find a near-optimal $\widetilde{z}_t$ such that $\|\widetilde{z}_t\|_0 \leq m$. The optimization for $\widetilde{z}_t$ will have the form

$$\widetilde{z}_t = \text{argmin}_{z \in \triangle_{\mathcal{H}}} \sum_{\tau=1}^{t} \nabla f_\tau(z_\tau)^\top z + 2\sigma_t (z_t - z_1)^\top z$$

for

$$
\begin{aligned}
[\nabla f_\tau(z_\tau)]_h &= c_h(\widetilde{\lambda}^1_\tau, \beta_\tau) \\
&= \min_{\widetilde{\lambda}^{2h} \in \triangle_n} \sum_j a_j \mathcal{I}\{h(u_j) \neq o_j\} + \sum_{i \in [n]} c_i \Big(\epsilon_k + \beta^h_{1\tau} \widehat{\Delta}^{k-1}(h) - \sum_{j \in [n]} \widetilde{\lambda}^{1h}_{j\tau} \beta^{hj}_{2\tau} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) \\
&\quad + \sum_{j \in [n]} \widetilde{\lambda}^{2h}_j \beta^{hj}_{3\tau} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big).
\end{aligned}
$$

Let $C_t \subseteq \mathcal{H}$ denote the classifiers supported on z_t and assume that z_1 is only supported on a single classifier h_0. Note from our discussion on computing the best-response for the $\widetilde{\lambda}^1$ and β player, we have that β^h and $\widetilde{\lambda}^{1h}$ are identical for all $h \notin C_t$. We can therefore rewrite the above objective as (dropping the τ subscript and denoting, e.g. $\beta^h_1 = \sum_{\tau=1}^t \beta^h_{1\tau}$):

$$
\begin{aligned}
&\min_{\widetilde{\lambda}^{2h} \in \triangle_n, h \in \mathcal{H}} \sum_{h \in \mathcal{H} \setminus C_t} z_h \Bigg(\sum_j a_j \mathcal{I}\{h(u_j) \neq o_j\} + \sum_{i \in [n]} c_i \Big(\epsilon_k + \beta_1 \widehat{\Delta}^{k-1}(h) - \sum_{j \in [n]} \widetilde{\lambda}^1_j \beta^j_2 \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) \\
&\quad + \sum_{j \in [n]} \widetilde{\lambda}^{2h}_j \beta^j_3 \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big) \Bigg) \\
&\quad + \sum_{h \in C_t} \Bigg(\sum_j a_j \mathcal{I}\{h(u_j) \neq o_j\} + \sum_{i \in [n]} c_i \Big(\epsilon_k + \beta^h_{1\tau} \widehat{\Delta}^{k-1}(h) - \sum_{j \in [n]} \widetilde{\lambda}^{1h}_{j\tau} \beta^{hj}_{2\tau} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) \\
&\quad + \sum_{j \in [n]} \widetilde{\lambda}^{2h}_j \beta^{hj}_{3\tau} \widehat{\Delta}^{j,k-1}_{\text{safe}}(h) - \widehat{\Delta}^{i,k}_{\text{safe}}(h) \Big) + 2\sigma_t z_t \Bigg) - 2\sigma_t z_{h_0}.
\end{aligned}
$$

We will focus first on the sum over $\mathcal{H} \setminus C_t$. Note that $\widehat{\Delta}^{k-1}(h)$ and $\widehat{\Delta}^{j,k}_{\text{safe}}(h)$ are both of the form

$$
\min_{\alpha \in \mathcal{A}} \sum_t \frac{1}{w_t + \alpha} \mathcal{I}\{h(u_t) \neq o_t\} + \alpha \sum_t \widetilde{w}_t \mathcal{I}\{h(u_t) \neq o_t\} + \frac{c}{\alpha}.
$$

Given this, we can rewrite the minimization over the first term as (where the $\widetilde{\alpha}$ correspond to the gaps that have negative coefficients, which is where the max comes from):

$$
\min_{z \in \triangle_{\mathcal{H}}} \min_{\widetilde{\lambda}^{2h} \in \triangle_n, h \in \mathcal{H}} \min_{\alpha^h \in \mathcal{A}^k, h \in \mathcal{H}} \max_{\widetilde{\alpha}^h \in \mathcal{A}^k, h \in \mathcal{H}} \sum_{h \in \mathcal{H} \setminus C_t} z_h \Big(\mathcal{R}(h; \alpha^h, \widetilde{\alpha}^h, \widetilde{\lambda}^{2h}) + f(\alpha^h, \widetilde{\lambda}^{2h}) + g(\widetilde{\alpha}^h, \widetilde{\lambda}^{2h}) \Big)
$$

for $\mathcal{R}$ convex in α, and concave in $\widetilde{\alpha}$, f convex in α, and g concave in $\widetilde{\alpha}$, and all functions are linear in $\widetilde{\lambda}^2$. Normally $\mathcal{A}$ is a discrete set, but if we let $\widetilde{\mathcal{A}}$ be a continuous relaxation of it, we can rewrite the above as

$$
\min_{z \in \triangle_{\mathcal{H}}} \max_{\widetilde{\alpha}^h \in \mathcal{A}^k, h \in \mathcal{H}} \min_{\widetilde{\lambda}^{2h} \in \triangle_n, h \in \mathcal{H}} \min_{\alpha^h \in \mathcal{A}^k, h \in \mathcal{H}} \sum_{h \in \mathcal{H} \setminus C_t} z_h \Big(\mathcal{R}(h; \alpha^h, \widetilde{\alpha}^h, \widetilde{\lambda}^{2h}) + f(\alpha^h, \widetilde{\lambda}^{2h}) + g(\widetilde{\alpha}^h, \widetilde{\lambda}^{2h}) \Big).
$$

To solve this we can again apply the FW algorithm of (Hazan and Kale, 2012) with the max player playing best-response. As before, as long as $\mathfrak{z}_t$ (where $\mathfrak{z}_t$ denotes the update for this inner optimization) is sparse, we can efficiently compute the best-response for the $\widetilde{\alpha}$ player, since we only need to compute it for $h \in \mathfrak{z}_t$. The FW-style update will then have the form

$$
\min_{z \in \triangle_{\mathcal{H}}} \min_{\widetilde{\lambda}^{2h} \in \triangle_n, h \in \mathcal{H}} \min_{\alpha^h \in \mathcal{A}^k, h \in \mathcal{H}} \sum_{h \in \mathcal{H} \setminus C_t} z_h \Big(\mathcal{R}(h; \alpha^h, \widetilde{\alpha}^h_t, \widetilde{\lambda}^{2h}) + f(\alpha^h, \widetilde{\lambda}^{2h}) + g(\widetilde{\alpha}^h_t, \widetilde{\lambda}^{2h}) \Big)
$$

$$= \min_{\widetilde{\lambda}^{2h} \in \triangle_n, h \in \mathcal{H}} \min_{\alpha^h \in \mathcal{A}^k, h \in \mathcal{H}} \min_{h \in \mathcal{H} \backslash C_t} \mathcal{R}(h; \alpha^h, \widetilde{\alpha}_t^h, \widetilde{\lambda}^{2h}) + f(\alpha^h, \widetilde{\lambda}^{2h}) + g(\widetilde{\alpha}_t^h, \widetilde{\lambda}^{2h})$$

where the equality follows since we can always swap min, and since there will always be an optimal solution supported on a single h. We can solve the inner min using a CSC oracle that is able to optimize over a set $\mathcal{H} \backslash C_t$, and by enumerating $\widetilde{\lambda}^2$ and α (since we can always find an optimal solution supported on a single h, we can set $\widetilde{\lambda}^{2h}, \alpha^h$ identical for all h and will arrive at the same minimum).

This will converge in polynomially many steps, and will produce some $\mathfrak{z}_{t'}$ which is m-sparse (for m polynomial in parameters). It follows that $\mathfrak{z}_{t'}$ is the near-optimal value for $\widetilde{z}_t$ supported on $\mathcal{H} \backslash C_t$. To pick a final value for $\widetilde{z}_t$, we can simply enumerate over the (polynomially many) $h \in C_t$, compute their loss values, and then pick the minimum out of those and the value achieved by $\mathfrak{z}_{t'}$. This procedure will always return some $\widetilde{z}_t$ supported on at most polynomially many h, so m can be chosen suitably to make the best-response of the max player efficient.

Putting all of this together, we can efficiently solve for $\widehat{y}_\ell$.

Solving for λ_ℓ

We turn now to solving the optimization (E.16). Using arguments similar to what we have already shown, we have that

$$\text{(E.16)} = \inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} \min_{\alpha \in \mathcal{A}, \alpha_2, \ldots, \alpha_p \in \mathcal{A}} \max_{\beta_1, \ldots, \beta_m \in \mathcal{B}} \mathcal{R}(z; \alpha, \alpha_2, \ldots, \alpha_p, \beta_1, \ldots, \beta_m)$$
$$+ 2\alpha \sum_{U \in \mathcal{X}} \frac{\nu(U)^2}{9\lambda(U)/10 + 1/10d} \mathcal{I}\{z(U) \neq \widehat{z}_{\ell-1}(U)\} + f(\alpha, \alpha_2, \ldots, \alpha_p, \beta_1, \ldots, \beta_m).$$

As before, we can simply enumerate over all possible choices of α and β. For a fixed setting of α and β, to solving the $\inf$ over λ, we can apply Mirror Descent. In this case we choose the mirror map to be the negative entropy, which is strongly convex with respect to the ℓ_1 norm.

Given this, to solve this in a computationally efficient manner, all we need is that the objective is convex (which it is) and Lipschitz with respect to the ℓ_1 norm. Let $g(\lambda)$ denote the objective of the above optimization. By the Mean Value Theorem,

$$|g(\lambda) - g(\widetilde{\lambda})| = \nabla g((1-c)\lambda + c\widetilde{\lambda})^\top (\lambda - \widetilde{\lambda})$$

for some $c \in [0, 1]$. So, for any $\lambda, \widetilde{\lambda} \in \triangle_{\mathcal{X}}$, we can bound

$$|g(\lambda) - g(\widetilde{\lambda})| \leq \left(\sup_{\lambda' \in \triangle_{\mathcal{X}}} \|\nabla g(\lambda')\|_\infty \right) \cdot \|\lambda - \widetilde{\lambda}\|_1.$$

We have,

$$\frac{d}{dt} \sum_{U \in \mathcal{X}} \frac{\nu(U)^2}{9\lambda(U)/10 + 1/10d + 9t\lambda_0(U)/10} \mathcal{I}\{z(U) \neq \widehat{z}_{\ell-1}(U)\}|_{t=0}$$
$$= \sum_{U \in \mathcal{X}} \frac{-\lambda_0(U)\nu(U)^2}{(9\lambda(U)/10 + 1/10d)^2} \mathcal{I}\{z(U) \neq \widehat{z}_{\ell-1}(U)\}.$$

It follows that

$$\sup_{\lambda' \in \triangle_{\mathcal{X}}} \|\nabla g(\lambda')\|_\infty \leq 100d^2$$

so we can apply Mirror Descent to optimize the above with computational complexity scaling only polynomially in problem parameters.

E.5.2 Computational Efficiency of BESIDE

The primary computational cost of BESIDE is incurred by calling RAGE$^\epsilon$, and solving the optimization on Line 6 of Algorithm 6.1. We have already shown that RAGE$^\epsilon$ can be run in a computationally efficient manner. The optimization on Line 6 has a form very similar to the optimization we solve in RAGE$^\epsilon$, so the same argument and solution approach (applying Mirror Descent) allows us to compute the optimal distribution, λ_ℓ, here as well.

E.6 Experimental details and additional results

E.6.1 Experimental details

All code was written in Python and run on a Intel Xeon 6226R CPU with 64 cores.

Algorithm E.4 is the precise implementation of BESIDE using elimination. It largely resemble to Algorithm 6.1, with the difference that it explicitly eliminates arms.

Algorithm E.3. Best Safe Arm Identification (BESIDE, defined with generic constants)

input: tolerance ϵ, confidence δ

$\iota_\epsilon \leftarrow \lceil \log(\frac{2}{\min\{c_3,c_4\}\cdot\epsilon}) \rceil$, $\widehat{\Delta}^0_{\text{safe}}(z) \leftarrow 0, \widehat{\Delta}^0(z) \leftarrow 0$ for all $z \in \mathcal{Z}$

for $\ell = 1, 2, \ldots, \iota_\epsilon$ **do**

$\epsilon_\ell \leftarrow \frac{2}{\min\{c_3,c_4\}} \cdot 2^{-\ell}$

// Solve experiment to reduce uncertainty on safety constraints

Let τ_ℓ be the minimal value of $\tau = 2^j \geq 4 \log \frac{4m|\mathcal{Z}|\ell^2}{\delta}$ such that the objective to the following is no greater than $c_e \epsilon_\ell$, and λ_ℓ the corresponding optimal distribution

$$\inf_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z}} -c_d \left(\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) + \epsilon_\ell \right) + \sqrt{\frac{\|z\|^2_{A(\lambda)^{-1}} \cdot \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}{\tau}}$$

Sample $x_t \sim \lambda_\ell$, collect τ_ℓ observations $\{(x_t, r_t, s_{t,1}, \ldots, s_{t,m})\}_{t=1}^{\tau_\ell}$

$\{\widehat{\mu}^{i,\ell}\}_{i=1}^m \leftarrow \mathsf{RIPS}(\{(x_t, s_{t,i})\}_{t=1}^{\tau_\ell}, \mathcal{Z}, \frac{\delta}{2m\ell^2})$ // Estimate safety constraints

$\widehat{\Delta}^{i,\ell}_{\text{safe}}(z) \leftarrow \gamma - z^\top \widehat{\mu}^{i,\ell} + \|z\|_{A(\lambda_\ell)^{-1}} \sqrt{\tau_\ell^{-1} \log(\frac{4m|\mathcal{Z}|\ell^2}{\delta})}$ // Safety gap estimates

// Form set of arms guaranteed to be safe

$$\mathcal{Y}_\ell \leftarrow \Big\{ z \in \mathcal{Z} : 8c_d \Big(\min_j |\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell-1}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\ell-1}(z)) \Big) + 8(c_d + c_e)\epsilon_\ell \leq \widehat{\Delta}^{i,\ell}_{\text{safe}}(z), \forall i \in [n] \Big\} \cup \mathcal{Y}_{\ell-1}$$

// Refine estimates of optimality gaps

$\{\widehat{\Delta}^\ell(z)\}_{z\in\mathcal{Z}} \leftarrow \text{RAGE}^\epsilon\Big(\mathcal{Z}, \mathcal{Y}_\ell, \epsilon_\ell, \frac{\delta}{4\ell^2}, \{\widehat{\Delta}_{\text{safe}}(z) \leftarrow \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\ell}_{\text{safe}}(z))\}_{z\in\mathcal{Z}}\Big)$

// Form set of arms guaranteed to be at most ϵ-unsafe

$$\mathcal{Y}_{\text{end}} \leftarrow \Big\{ z \in \mathcal{Z} : 8c_d \Big(\min_j |\widehat{\Delta}^{j,\iota_\epsilon}_{\text{safe}}(z)| + \max_j \mathfrak{p}(-\widehat{\Delta}^{j,\iota_\epsilon}_{\text{safe}}(z)) + \mathfrak{p}(\widehat{\Delta}^{\iota_\epsilon}(z)) \Big) + 8(c_d + c_e)\epsilon - c_g\epsilon \leq \widehat{\Delta}^{i,\iota_\epsilon}_{\text{safe}}(z), \forall i \in [n] \Big\}$$

// Find ϵ-good arm out of ϵ-safe arms

$\{\widehat{\Delta}^{\text{end}}(z)\}_{z\in\mathcal{Y}_{\text{end}}} \leftarrow \text{RAGE}^\epsilon(\mathcal{Y}_{\text{end}}, \mathcal{Y}_{\text{end}}, \epsilon, \delta)$

return $\widehat{z} = \text{argmin}_{z\in\mathcal{Y}_{\text{end}}} \widehat{\Delta}^{\text{end}}(z)$

Algorithm E.4. Best Safe Arm Identification with Elimination

input: tolerance ϵ, confidence δ
$\iota_\epsilon \leftarrow \lceil \log(\frac{1}{\epsilon}) \rceil$, $\mathcal{Z}^0_{\text{active}} \leftarrow \mathcal{Z}$, $\mathcal{Z}^0_{\text{safe}} \leftarrow \emptyset$
for $\ell = 1, 2, \ldots, \iota_\epsilon$ **do**
    $\epsilon_\ell \leftarrow 2^{-\ell}$
    Compute allocation $\mathcal{XY}_{\text{safe}}$ on $\mathcal{Z}^{\ell-1}_{\text{active}}$ and sample from it $\tau_\ell = \mathcal{O}(\mathcal{XY}_{\text{safe}}(\mathcal{Z}^{\ell-1}_{\text{active}})/\epsilon_\ell^2)$ times
    $\widehat{\mu}^\ell \leftarrow \mathsf{RIPS}(\{(x_t, s_{t,i})\}_{t=1}^{\tau_\ell}, \mathcal{Z}, \frac{\delta}{2\ell^2})$
    Set $\widehat{\Delta}^\ell_{\text{safe}}(z) \leftarrow \gamma - z^\top \widehat{\mu}^\ell$ for all $z \in \mathcal{Z}^{\ell-1}_{\text{active}}$ and

$$\widetilde{Z}^\ell_{\text{active}} = \{z \in \widetilde{Z}^{\ell-1}_{\text{active}} \ : \ \widehat{\Delta}^\ell_{\text{safe}}(z) \in [-\epsilon_\ell, 2\epsilon_\ell]\} \quad \widetilde{Z}^\ell_{\text{safe}} = \{z \in \widetilde{Z}^{\ell-1}_{\text{active}} \ : \ \widehat{\Delta}^\ell_{\text{safe}}(z) \geq 2\epsilon_\ell]\}$$

    $\mathcal{Z}^\ell_{\text{active}}, \mathcal{Z}^\ell_{\text{safe}} \leftarrow \text{RAGE-ELIM}^\epsilon\Big(\widetilde{Z}^\ell_{\text{active}} \cup \widetilde{Z}^\ell_{\text{safe}} \cup \mathcal{Z}^{\ell-1}_{\text{safe}}, \widetilde{Z}^\ell_{\text{safe}} \cup \mathcal{Z}^{\ell-1}_{\text{safe}}, \epsilon_\ell\Big)$
$\mathcal{Z}_{\text{final}}, \emptyset \leftarrow \text{RAGE-ELIM}^\epsilon\Big(\mathcal{Z}^\ell_{\text{active}} \cup \mathcal{Z}^\ell_{\text{safe}}, \mathcal{Z}^\ell_{\text{active}} \cup \mathcal{Z}^\ell_{\text{safe}}, \epsilon_\ell\Big)$
return Any arm in $\mathcal{Z}_{\text{final}}$.

Algorithm E.5. RAGE-ELIM$^\epsilon$

input: active set $\mathcal{Z}$, optimal set $\mathcal{Y}$, tolerance ϵ
$\iota_\epsilon \leftarrow \lceil \log(\frac{1}{\epsilon}) \rceil$, $\mathcal{Z}^0 \leftarrow \mathcal{Z}$, $\mathcal{Y}^0 \leftarrow \mathcal{Y}$
for $\ell = 1, 2, \ldots, \iota_\epsilon$ **do**
    $\epsilon_\ell \leftarrow 2^{-\ell}$
    Compute allocation $\mathcal{XY}_{\text{diff}}$ on $(\mathcal{Z}^{\ell-1} \cup \mathcal{Y}^{\ell-1}, \mathcal{Y}^{\ell-1})$ and sample from it $\tau_\ell = \mathcal{O}(\mathcal{Z}^{\ell-1} \cup \mathcal{Y}^{\ell-1}, \mathcal{Y}^{\ell-1})/\epsilon_\ell^2)$ times
    $\widehat{\theta}^\ell \leftarrow \mathsf{RIPS}(\{(x_t, s_{t,i})\}_{t=1}^{\tau_\ell}, \mathcal{Z}, \frac{\delta}{2\ell^2})$
    Set $\widehat{\Delta}^\ell(z) \leftarrow \max_{y \in \mathcal{Y}^{\ell-1}} y^\top \widehat{\theta}^\ell - z^\top \widehat{\theta}^\ell$ for all $z \in \mathcal{Z} \cup \mathcal{Y}$ and

$$\mathcal{Z}^\ell = \{z \in \mathcal{Z}^{\ell-1} \ : \ \widehat{\Delta}^\ell(z) \leq \epsilon_\ell\} \quad \mathcal{Y}^\ell = \{y \in \mathcal{Y}^{\ell-1} \ : \ \widehat{\Delta}^\ell(y) \leq \epsilon_\ell\}$$

return $\mathcal{Z}^\ell, \mathcal{Y}^\ell$

Appendix F

Appendix for Chap. 7

F.1 Datasets description

***`Adult income` dataset* (Lichman, 2013)**: This dataset comprises $48,842$ examples with demographic information. The task is to predict whether an individual's income exceeds $50k\$$ annually. We chose the protected attribute to be binarized gender.

***`Compas` dataset* (Lichman, 2013)**: This dataset, which was released by Angwin et al. (2022), encompasses $5,278$ data related to juvenile felonies. It includes details such as marital status, ethnicity, age, prior criminal history, and the severity of the current arrest charges. In our analysis, we identify binarized gender as a sensitive attribute. In line with established conventions (Corbett-Davies et al., 2017; Anahideh et al., 2021), we adopt a two-year violent recidivism record as the ground truth for assessing recidivism.

`Drug consumption` dataset (Fehrman et al., 2017): This dataset consists of $1,885$ entries containing information about individuals, where each entry includes five demographic characteristics (such as Age, binarized Gender, or Education), seven measurements related to personality traits (such as Nscore indicating neuroticism and Ascore representing agreeableness), and 18 descriptors detailing the subject's most recent consumption of a specific substance (like Cannabis). We chose the task of predicting whether an individual consumed Cannabis in the last year and chose the protected attribute to be (binarized) Gender.

`German Credit` dataset (Hofmann, 1994): The German Credit dataset classifies people as good or bad credit risks using the profile and history of $1,000$ clients. We set the binarized gender as the sensitive attribute.

`Community and Crime` dataset (Redmond and Baveja, 2002): The Crime and Community dataset consists of $1,902$ instances of crimes with 128 attributes related to the crime and the corresponding community. It uses 'violent crimes' as the target variable and combines 'percentage of non-white' as the protected attribute. The target variable is binarized to categorize communities as high or low crime based on a threshold of 500. The protected attribute is also binarized, separating communities with non-white residents below 20%.

`Bank` dataset (Moro et al., 2014): The task is to predict whether the client has subscribed to a term deposit service based on $11,162$ data points with features such as marital status and age. We set the client having tertiary education as the sensitive attribute.

`Synthetic` dataset: We created the synthetic dataset in the following manner. It is depicted in Figure F.1. The dataset consists of two dimensions, and data for group 0 is generated by randomly sampling $10,000$ data points from a Gaussian distribution with a mean of $(0, 0)$, while group 1 comprises 100 data points sampled from $(10, 10)$. For group 0 (and group 1), labels are assigned a value of 1 if the x-coordinate

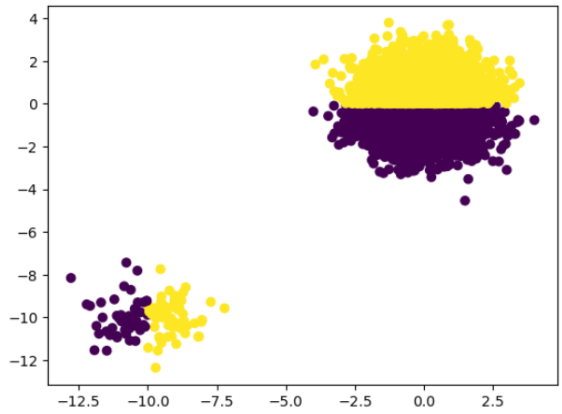


Figure F.1: Synthetic dataset

(or y-coordinate) of the data point is greater than 0, and 0 otherwise. This ensures that each group is linearly separable, but their combination is not.

F.2 Performance of baseline algorithms with different pre-trained dataset sizes

We report the results of the sweeps over the size of the pretrain dataset in Figures F.2 to F.13. Due to its large computational cost, we compared the performance of PANDA for two sizes of pretrain datasets.

F.3 Theoretical results - proof of Proposition 2

F.3.1 Full theorem

We have the following result.

Theorem 32. *Let the train set be* $\mathcal{D} = \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\}$. *If* $\mathcal{D} \sim \nu$, *then it holds with probability* $1 - \delta$ *that:*

$$
\begin{aligned}
|L_\nu^{\mathrm{EO}}(h) - \widehat{L}_\mathcal{D}^{\mathrm{EO}}(h)| &\leq C_{0,0} + C_{0,1} + C_{1,0} + C_{1,1}, \\
|L_\nu^{\mathrm{TP}}(h) - \widehat{L}_\mathcal{D}^{\mathrm{TP}}(h)| &\leq C_{0,1} + C_{1,1}, \\
|L_\nu^{\mathrm{FP}}(h) - \widehat{L}_\mathcal{D}^{\mathrm{FP}}(h)| &\leq C_{0,0} + C_{1,0},
\end{aligned}
$$

with confidence terms

$$
\begin{aligned}
C_{j,k} = \left(\widehat{p}_{j,k} + \sqrt{2\widehat{\mathcal{V}}_{j,k}^{(1)} \frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n} \right) \cdot \frac{\sqrt{2\widehat{\mathcal{V}}_{j,k}^{(2)} \frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{y_i = k, a_i = j\}\right)^2} \\
+ \frac{\sqrt{2\widehat{\mathcal{V}}_{j,k}^{(1)} \frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{y_i = k, a_i = j\}}
\end{aligned}
$$

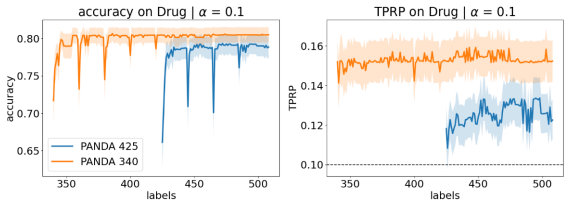


Figure F.2: Performance on `Drug Consumption`

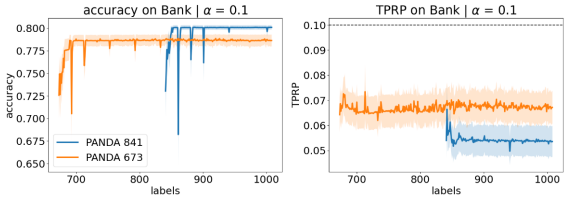


Figure F.3: Performance on `Bank`

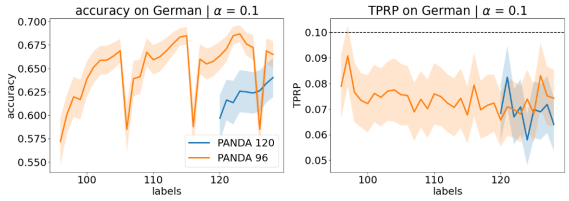


Figure F.4: Performance on `German Credit`

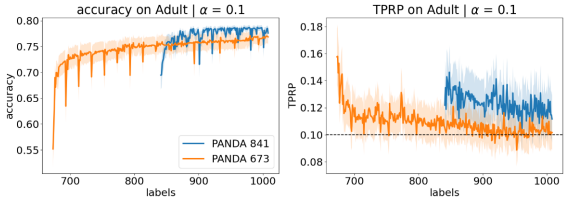


Figure F.5: Performance on `Adult Income`

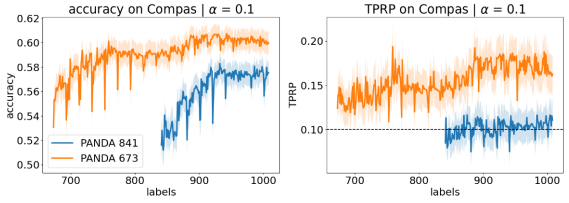


Figure F.6: Performance on `Compas`

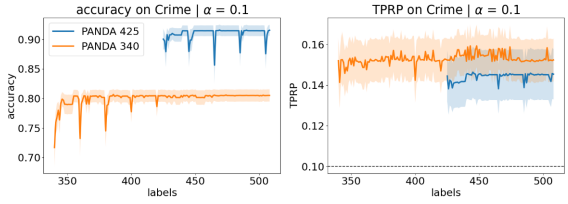


Figure F.7: Performance on `Community and Crime`

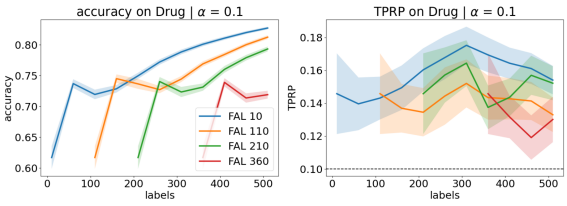


Figure F.8: Performance on `Drug Consumption`

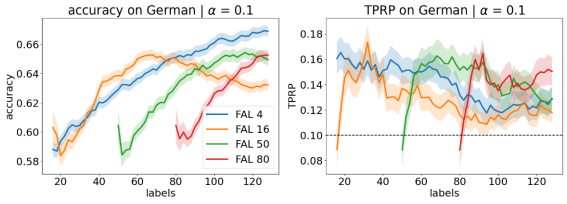


Figure F.10: Performance on `German Credit`

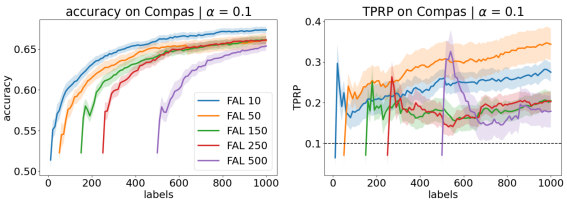


Figure F.12: Performance on `Compas`

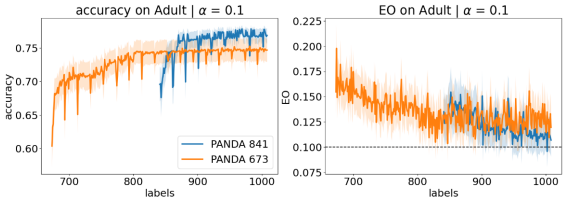


Figure F.14: Performance on `Adult Income` for Equalized Odds

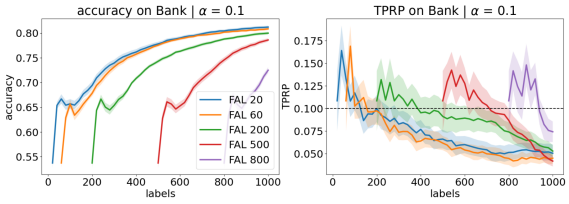


Figure F.9: Performance on `Bank`

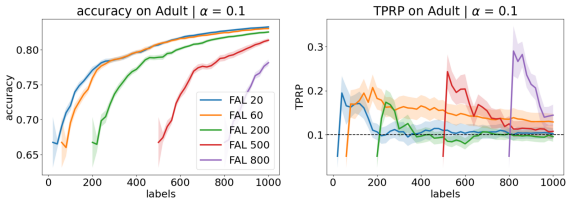


Figure F.11: Performance on `Adult Income`

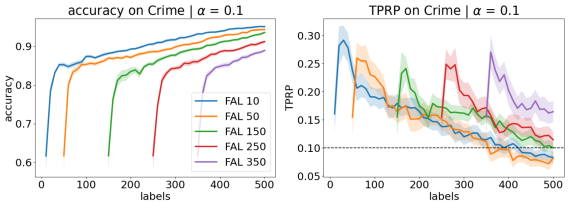


Figure F.13: Performance on `Community and Crime`

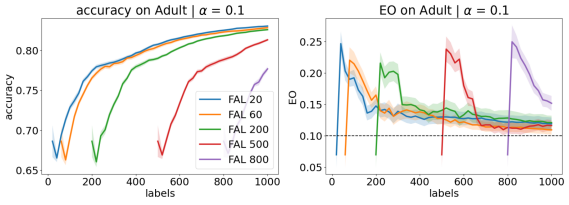


Figure F.15: Performance on `Adult Income` for Equalized Odds

for label $k \in \{0,1\}$ and protected attribute $j \in \{0,1\}$, where $\widehat{p}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{h(x_i)=1, y_i = k, a_i = j\}$ and the empirical variances defined as

$$\widehat{\mathcal{V}}_{j,k}^{(1)} = \frac{1}{n(n-1)} \sum_{1 \le \ell < \ell' \le n} (\mathbf{1}\{h(x_\ell)=1, y_\ell = k, a_\ell = j\} - \mathbf{1}\{h(x_{\ell'})=1, y_{\ell'} = k, a_{\ell'} = j\})^2,$$
$$\widehat{\mathcal{V}}_{j,k}^{(2)} = \frac{1}{n(n-1)} \sum_{1 \le \ell < \ell' \le n} (\mathbf{1}\{y_\ell = k, a_\ell = j\} - \mathbf{1}\{y_{\ell'} = k, a_{\ell'} = j\})^2.$$

This theorem provides a confidence bound on the concentration rate of the empirical fairness violation.

Proof. Let us start by proving the statement for TPRP. Recall

$$L_\nu^{\mathrm{TP}}(h) = \left| \frac{P_{(x,a,y)\sim\nu}(h(x)=1, a=0, y=1)}{P_{(x,a,y)\sim\nu}(a=0,y=1)} - \frac{P_{(x,a,y)\sim\nu}(h(x)=1, a=1, y=1)}{P_{(x,a,y)\sim\nu}(a=1,y=1)} \right|$$
$$\widehat{L}_{\mathcal{D}}^{\mathrm{TP}}(h) = \left| \sum_{i=1}^{n} \frac{\mathbf{1}\{h(x_i)=1, y_i=1, a_i=1\}}{\sum_{i=1}^{n}\mathbf{1}\{y_i=1,a_i=1\}} - \sum_{i=1}^{n} \frac{\mathbf{1}\{h(x_i)=1, y_i=1, a_i=0\}}{\sum_{i=1}^{n}\mathbf{1}\{y_i=1,a_i=0\}} \right|.$$

and write these for short

$$L_\nu^{\mathrm{TP}}(h) = |\mathrm{num}_0/\mathrm{den}_0 - \mathrm{num}_1/\mathrm{den}_1|,$$
$$\widehat{L}_{\mathcal{D}}^{\mathrm{TP}}(h) = |\widehat{\mathrm{num}}_0/\widehat{\mathrm{den}}_0 - \widehat{\mathrm{num}}_1/\widehat{\mathrm{den}}_1|,$$

with for protected attribute $j \in \{0,1\}$,

$$\mathrm{num}_j = P_{(x,a,y)\sim\nu}(h(x)=1, a=j, y=1)$$
$$\widehat{\mathrm{num}}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{h(x_i)=1, y_i=1, a_i=j\}$$
$$\mathrm{den}_j = P_{(x,a,y)\sim\nu}(a=j, y=1)$$
$$\widehat{\mathrm{den}}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i=1, a_i=j\}.$$

Applying Bernstein's concentration bound it holds that for $j \in \{0,1\}$ with probability at least $1-\delta$

$$|\widehat{\mathrm{num}}_j - \mathrm{num}_j| = \left| \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{h(x_i)=1, y_i=1, a_i=j\} - \mathbb{P}_{(x,a,y)\sim\nu}(h(x)=1, y=1, a=j) \right|$$
$$\le \sqrt{2\widehat{\mathcal{V}}_{j,1}^{(1)} \frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n} =: \alpha_j^{(\mathrm{num})},$$

where we defined

$$\widehat{\mathcal{V}}_{j,k}^{(1)} = \frac{1}{n(n-1)} \sum_{1 \le \ell < \ell' \le n} (\mathbf{1}\{h(x_\ell)=1, y_\ell = k, a_\ell = j\} - \mathbf{1}\{h(x_{\ell'})=1, y_{\ell'} = k, a_{\ell'} = j\})^2.$$

Also applying Bernstein's concentration bound it holds that for $j \in \{0,1\}$ with probability at least $1-\delta$

$$|\widehat{\mathrm{den}}_j - \mathrm{den}_j| = \left| \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i=1, a_i=j\} - \mathbb{P}_{(x,a,y)\sim\nu}(y=1, a=j) \right|$$

$$\leq \sqrt{2\widehat{\mathcal{V}}_{j,1}^{(2)}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n} =: \alpha_j^{(\text{den})},$$

where we defined

$$\widehat{\mathcal{V}}_{j,k}^{(2)} = \frac{1}{n(n-1)} \sum_{1\leq \ell < \ell' \leq n} (\mathbf{1}\{y_\ell = k, a_\ell = j\} - \mathbf{1}\{y_{\ell'} = k, a_{\ell'} = j\})^2.$$

Then, as soon as for both $j = 1$ and $j = 2$, $\alpha_j^{(\text{den})} \leq \widehat{\text{den}}_j/2$, holds the inequality

$$\left| \frac{1}{\widehat{\text{den}}_j} - \frac{1}{\text{den}_j} \right| \leq \frac{\alpha_j^{(\text{den})}}{\widehat{\text{den}}_j^{\,2}},$$

so that for $j \in \{0, 1\}$, we have

$$\begin{aligned}
\left| \frac{\widehat{\text{num}}_j}{\widehat{\text{den}}_j} - \frac{\text{num}_j}{\text{den}_j} \right| &= \left| \frac{\widehat{\text{num}}_j}{\widehat{\text{den}}_j} - \frac{\text{num}_j}{\widehat{\text{den}}_j} - \frac{\text{num}_j}{\widehat{\text{den}}_j} - \frac{\text{num}_j}{\text{den}_j} \right| \\
&\leq \left| \frac{\widehat{\text{num}}_j}{\widehat{\text{den}}_j} - \frac{\text{num}_j}{\widehat{\text{den}}_j} \right| + \left| \frac{\text{num}_j}{\widehat{\text{den}}_j} - \frac{\text{num}_j}{\text{den}_j} \right| \\
&\leq \frac{\alpha_j^{(\text{num})}}{\widehat{\text{den}}_j} + \frac{\text{num}_j \alpha_j^{(\text{den})}}{\widehat{\text{den}}_j^{\,2}} \\
&\leq \frac{\alpha_j^{(\text{num})}}{\widehat{\text{den}}_j} + \frac{(\alpha_j^{(\text{num})} + \widehat{\text{num}}_j)\alpha_j^{(\text{den})}}{\widehat{\text{den}}_j^{\,2}}.
\end{aligned}$$

Note that $C_{j,1}$ is exactly the last upper bound above,

$$\begin{aligned}
C_{j,1} = &\left(\widehat{p}_{j,1} + \sqrt{2\widehat{\mathcal{V}}_{j,1}^{(1)}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n} \right) \cdot \frac{\sqrt{2\widehat{\mathcal{V}}_{j,1}^{(2)}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{y_i = 1, a_i = j\}\right)^2} \\
&+ \frac{\sqrt{2\widehat{\mathcal{V}}_{j,1}^{(1)}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{y_i = 1, a_i = j\}}
\end{aligned}$$

where $\widehat{p}_{j,1} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{h(x_i) = 1, y_i = 1, a_i = j\}$. Putting it together

$$\begin{aligned}
|L_\nu^{\text{TP}}(h) - \widehat{L}_{\mathcal{D}}^{\text{TP}}(h)| &= ||\text{num}_0/\text{den}_0 - \text{num}_1/\text{den}_1| - |\widehat{\text{num}}_0/\widehat{\text{den}}_0 - \widehat{\text{num}}_1/\widehat{\text{den}}_1||, \\
&\leq |\text{num}_0/\text{den}_0 - \text{num}_1/\text{den}_1 - \widehat{\text{num}}_0/\widehat{\text{den}}_0 + \widehat{\text{num}}_1/\widehat{\text{den}}_1|, \\
&\leq |\text{num}_0/\text{den}_0 - \widehat{\text{num}}_0/\widehat{\text{den}}_0| + |\widehat{\text{num}}_1/\widehat{\text{den}}_1 - \text{num}_1/\text{den}_1|, \\
&\leq C_{0,1} + C_{1,1}.
\end{aligned}$$

which is the conclusion for TPRP.

As $\widehat{L}^{\mathrm{FP}}_{\mathcal{D}}(h)$ was defined as the empirical estimate of the FPRP violation by conditioning on $\mathbf{1}\{y_i = 0\}$ (instead of $\mathbf{1}\{y_i = 1\}$ for TPRP), the proof for the concentration bound on FPRP is analogous to the one of TPRP, with the exception of the conditioning on $\mathbf{1}\{y_i = 0\}$ instead of $\mathbf{1}\{y_i = 1\}$ for TPRP.

We defined the empirical estimate of the EO violation as the maximum of empirical estimate of the TPRP violation and the empirical estimate of the FPRP violation, $\widehat{L}^{\mathrm{EO}}_{\mathcal{D}}(h) = \max\{\widehat{L}^{\mathrm{TP}}_{\mathcal{D}}(h), \widehat{L}^{\mathrm{FP}}_{\mathcal{D}}(h)\}$, so holds

$$\widehat{L}^{\mathrm{EO}}_{\mathcal{D}}(h) \leq \widehat{L}^{\mathrm{TP}}_{\mathcal{D}}(h) + \widehat{L}^{\mathrm{FP}}_{\mathcal{D}}(h),$$

which immediately leads to the conclusion of Theorem 32. □

F.3.2 Proof of Proposition 2

We first state the full result that leads to the statement of Proposition 2.

Proposition 8. *Let the train set be $\mathcal{D} = \{(x_1, a_1, y_1), \ldots, (x_n, a_n, y_n)\}$. If $\mathcal{D} \sim \nu$, then it holds with probability $1 - \delta$ that:*

$$|L^{\mathrm{TP}}_{\nu}(h) - \widehat{L}^{\mathrm{TP}}_{D}(h)| \leq 2 \max_{j \in \{0,1\}} \left\{ 2 \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = j\}} \right) + \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + 2\frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = j\}} \right)^2 \right\},$$

$$|L^{\mathrm{EO}}_{\nu}(h) - \widehat{L}^{\mathrm{EO}}_{D}(h)| \leq 4 \max_{0 \leq j,k \leq 1} \left\{ 2 \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \right) + \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + 2\frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \right)^2 \right\}.$$

Proof of Proposition 2 and 8. We use Theorem 32 and for label $k \in \{0, 1\}$ and protected attribute $j \in \{0, 1\}$ we bound $C_{j,k}$.

We first have that the empirical variances are such that $\widehat{\mathcal{V}}^{(1)}_{j,k} \leq 1$ and $\widehat{\mathcal{V}}^{(2)}_{j,k} \leq 1$. Also,

$$\begin{aligned}
\widehat{p}_{j,k} &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{h(x_i) = 1, y_i = k, a_i = j\} \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}.
\end{aligned}$$

Thus, we can bound

$$\begin{aligned}
C_{j,k} &= \left(\widehat{p}_{j,k} + \sqrt{2\widehat{\mathcal{V}}^{(1)}_{j,k}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n} \right) \cdot \frac{\sqrt{2\widehat{\mathcal{V}}^{(2)}_{j,k}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}\right)^2} \\
&\quad + \frac{\sqrt{2\widehat{\mathcal{V}}^{(1)}_{j,k}\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \\
&\leq 2 \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \right) + \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} 2\frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \right)^2.
\end{aligned}$$

With that result, we conclude for TPRP that

$$|L^{\mathrm{TP}}_{\nu}(h) - \widehat{L}^{\mathrm{TP}}_{D}(h)| \leq C_{0,1} + C_{1,1}$$

$$
\begin{aligned}
&\leq 2 \max_{j\in\{0,1\}} C_{j,1} \\
&\leq 2 \max_{j\in\{0,1\}} \left\{ 2 \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + \frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = j\}} \right) \right. \\
&\qquad \left. + \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + 2\frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = j\}} \right)^2 \right\} \\
&= 4 \max_{j\in\{0,1\}} \frac{\sqrt{2\frac{\log(2/\delta)}{n}}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = 1, a_i = j\}} + \mathcal{O}\left(\frac{1}{n}\right).
\end{aligned}
$$

Analogous bounds conclude for EO:

$$
\begin{aligned}
|L_\nu^{\mathrm{EO}}(h) - \widehat{L}_D^{\mathrm{EO}}(h)| \leq C_{0,0} + C_{1,0} + C_{0,1} + C_{1,1} &\leq 4 \max_{j\in\{0,1\}} C_{j,k} \\
&\leq 4 \max_{0\leq j,k\leq 1} \left\{ 2 \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}}\frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \right) \right. \\
&\qquad \left. + \left(\frac{\sqrt{2\frac{\log(2/\delta)}{n}} + 2\frac{\log(2/\delta)}{n}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} \right)^2 \right\} \\
&= 8 \max_{0\leq j,k\leq 1} \frac{\sqrt{2\frac{\log(2/\delta)}{n}}}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i = k, a_i = j\}} + \mathcal{O}\left(\frac{1}{n}\right).
\end{aligned}
$$

□