Data-Driven Sensor Placement Methods

Krithika Manohar

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

J. Nathan Kutz, Chair

Steven L. Brunton, Chair

Aleksandr Y. Aravkin

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

Data-Driven Sensor Placement Methods

Krithika Manohar

Co-Chairs of the Supervisory Committee:
Professor J. Nathan Kutz
Applied Mathematics

Professor Steven L. Brunton
Mechanical Engineering

The scalable optimization of sensor placement remains an open challenge in engineering and physical sciences. Optimal placements can only be determined in general using a brute-force combinatorial search over the domain. This explosion in complexity presents a major challenge for high-dimensional domains in oceanography, fluid dynamics, manufacturing, and biology. Fortunately, high-dimensional data generated by these systems often possess reproducible, low-rank structure that can be exploited to drastically reduce the amount of measurements required for global inference. In this thesis, we exploit data-driven learning to optimize sensor placement for signal reconstruction. Dimensionality reduction methods including proper orthogonal decomposition (POD) and dynamic mode decomposition (DMD) are used to obtain low-rank representations from data. We exploit empirical interpolation methods (EIMs), initially pioneered in reduced order modeling, to efficiently optimize the placement of sensors for high-dimensional reconstruction, estimation and control. This work connects our EIM-based method to related placement criteria in optimal experimental design, and extends our method to obtain an arbitrary number of optimal sensors. The superior performance and accuracy of our method is demonstrated on a variety of high-dimensional data from facial images, ocean temperatures, fluid dynamics, aircraft manufacturing and insect flight. Finally, an extension to sensor and actuator placement for optimal closed-loop control is proposed, which similarly leverages balanced model reduction of observability and controllability Gramians for speedy sensor and actuator optimization.

TABLE OF CONTENTS

Page

List of Figures ii

List of Tables iii

Chapter 1: Introduction 1
  1.1 Motivation 1
  1.2 Historical perspectives 3
  1.3 Sparsity and data-driven learning 4
  1.4 Organization and contributions 5
  1.5 Conventions and Notations 7
  1.6 Published work 9

Chapter 2: Sparsity and dimensionality reduction 10
  2.1 Compression in a generic dictionary 11
    2.1.1 Theory of compressed sensing 11
  2.2 Interpolation in a tailored dictionary 15
    2.2.1 Proper Orthogonal Decomposition 15
    2.2.2 Empirical Interpolation Methods 16

Chapter 3: Sensor placement for signal reconstruction 20
  3.1 Gappy POD reconstruction 21
  3.2 Optimal sensor placement objective 23
  3.3 QR pivoting method 24
  3.4 Connection to optimal experimental design 27
  3.5 Results and discussion 29
    3.5.1 Flow past a cylinder 30
    3.5.2 Noise comparison study 32

3.5.3 Extended Yale B eigenfaces 34
3.5.4 Sea surface temperature (SST) 37
3.6 Conclusions 39
3.6.1 Potential applied impact 41

Chapter 4: Sensors for predictive manufacturing 43
4.1 Motivation 43
4.2 Robust Principal Component Analysis 45
4.3 Measurement optimization and gap prediction 47
4.4 Results and discussion 48
4.4.1 Prediction results for entire wing to body join 49
4.4.2 Prediction results for segmented shims 51
4.5 Conclusions 54

Chapter 5: Optimized Sampling for multiscale dynamics 56
5.1 Background 57
5.1.1 Dynamic mode decomposition 57
5.1.2 Multiresolution DMD 60
5.2 Multiresolution analysis framework 62
5.2.1 Mathematical formulation 63
5.2.2 Online estimation and prediction 64
5.2.3 Offline measurement selection 66
5.3 Applications 67
5.3.1 Multiscale video example 68
5.3.2 NOAA ocean surface temperature 72
5.4 Conclusions 77

Chapter 6: Sparse approximation for insect flight dynamics 80
6.1 Motivation 80
6.2 Numerical Model of Fluid-Structure Interaction 82
6.3 Sparse frequency classification of wing strain 88
6.3.1 Sparse approximation for classification (SRC) 90
6.3.2 Optimal measurement selection for SRCf 91
6.4 Results and discussion 93

6.4.1 Feedback identification from full state 95
6.4.2 Sparse (ℓ_1) vs. least squares (ℓ_2) approximation 96
6.4.3 SRCf with QR selection 97
6.4.4 SRCf with spatial bias 99
6.5 Conclusions 101
6.5.1 Control Implications 101
6.5.2 Biological Implications 102

Chapter 7: Sensor and actuator placement for optimal control 103
7.1 Motivation 104
7.2 Background on balanced model reduction 105
7.2.1 Observability and controllability 105
7.2.2 Balanced truncation and the balancing transformation 107
7.3 Sensor and actuator placement using balanced truncation 110
7.3.1 Sensor placement 110
7.3.2 Actuator placement 114
7.4 Results and discussion 116
7.4.1 Random state space system 116
7.4.2 Linearized Ginzburg-Landau with stochastic disturbances 118
7.5 Conclusions 122

Chapter 8: Conclusion 124
8.1 Extension to nonlinear dynamics and inference 125

Bibliography 127

LIST OF FIGURES

Figure Number Page

1.1 Overview of data-driven sensor placement 6

2.1 Overview of compressed sensing 12
2.2 Compressed sensing, applied to three-tone signal. 14

3.1 Gappy POD 21
3.2 Flow past a cylinder 31
3.3 Singular values of cylinder flow 32
3.4 Comparison of QR pivoting with DEIM 33
3.5 Singular values of eigenfaces 35
3.6 Comparison of sensor placement methods 36
3.7 Oversampling with QR pivoting 37
3.8 Singular values of SST 38
3.9 QR pivoting sensors for reconstruction 39
3.10 Reconstructing sea surface temperatures 40

4.1 Overview of predictive shimming. 44
4.2 Robust principal components of shim gaps 47
4.3 Median gap prediction error 49
4.4 Pointwise absolute error distribution 50
4.5 Gap measurements 51
4.6 Optimal sensors for single segment 52
4.7 Pointwise error vs. number of optimal sensors 53
4.8 Comparison of absolute error distribution 53

5.1 Multiresolution analysis and sampling schematic 62
5.2 Multiscale video 69
5.3 Comparison of spatial modes and sensors 70
5.4 Comparison of estimated coefficients 71

5.5 Noise corrupted measurements 72
5.6 MrDMD amplitude map for sea surface temperature 73
5.7 MrDMD vs. POD sensors 74
5.8 ENSO coefficient prediction 76
5.9 Sensor Ensemble 77
5.10 Flow field reconstruction from sensors 78

6.1 Numerical model of insect wing 82
6.2 Strain dynamics 87
6.3 POD modes of strain dynamics 88
6.4 Frequency modes of strain dynamics 89
6.5 Sparse representation for classifying frequencies 90
6.6 Frequency measurement selection 92
6.7 ℓ_1 approximation performance 95
6.8 Comparison to least-squares approximation 97
6.9 Ensemble of QR frequency selections 98
6.10 SRCf accuracy 99
6.11 Chordwise classification accuracy 100

7.1 Illustration of balancing transformation on Gramians 108
7.2 Sensor and actuator placement in random state-space system. 117
7.3 LQG controller for linearized Ginzburg-Landau 119
7.4 Sensor+actuator placements for linearized Ginzburg-Landau 120
7.5 LQG gain for five-input, five-output system. 121

LIST OF TABLES

Table Number Page

1.1 Nomenclature 8

3.1 Summary of sensor placement methods 30

4.1 Gap prediction from measurements 52

5.1 Multiscale video dynamics 68

6.1 Additional notation 83

6.2 Summary of numerical experiments and sparse classifiers 94

ACKNOWLEDGMENTS

First and foremost, I thank my advisors, Nathan Kutz and Steve Brunton, who guided me through numerous exciting research topics, collaborations, and applications. Many have complained about being pulled in different directions by multiple advisors, but fortunately, I am not one of them. Nathan and Steve work as seamlessly with me as they do with each other, and my numerous interactions with them both as a team and individually have been invaluable. They lead by example, in their research, teaching, mentorship, and friendship. One can only hope that some of this has trickled down to me.

I thank my collaborators on this work, Eurika Kaiser, Bing Brunton, Tom Hogan, Jim Buttrick, and Ashis Banerjee, for their invaluable insights on topics ranging from sparse sensing to aircraft assembly. I also thank my committee members Sasha Aravkin and Jevin West for their timely questions and perspectives. I am forever grateful to my colleagues and groupmates who have deepened my understanding of math on far more occasions than I care to admit. My undergraduate advisors (in official and unofficial capacities), Ronald Brent, James Canning, Georges Grinstein and Flip Johnson, motivated me to study mathematics, and for this I owe them dearly.

The last five years in Seattle have been profound and life-changing, and not only academically. There are those who sustained my work, and those who have sustained me. Those friendships that showed me how little I knew outside of math, have given me happiness, comfort and love beyond comprehension. For the fondest of memories, I thank Kimberly Latham, Ahmed Salama, Angela Liu and Lily Jiang.

It goes without saying that I owe my family for everything that cannot be expressed in words. I thank my parents, Sadhana and Sanjeev Manohar, for giving me the means, motivation and freedom to pursue my interests. My mother is my role model for success, for she is the strongest, most brilliant person I know. My father, my perpetual advisor, once told me it is never too late to switch tracks and follow any dream, and he meant it. Neha, my sister, has been my constant throughout childhood, and I have missed her tremendously these last five years. All my love goes to them, and to Venki Appa, who is always with us.

The work presented in this thesis was supported by the Boeing Corporation (grant SSOW-BRT-W0714-0004), Army Research Office (grants W911NF-17-1-0306 and W911NF-17-1-0422), the Air Force Office of Scientific Research (grant AFOSR FA9550-15-1-0385), and a Seattle ARCS Foundation Fellowship.

DEDICATION

to my family

Chapter 1

INTRODUCTION

Optimal sensor placement determines where to place sensors in a system to optimize specific downstream objectives. The topic, also known as optimal experimental design and variable selection, spans multiple disciplines and guides efficient decision-making in robotics, navigation, ocean monitoring, biological responses and more. Although sensors themselves take many different forms, they share a common interpretation - sensors provide local observables from which global properties may be inferred, and they are selected from a larger set of candidate locations. However, existing sensor placement methods are not scalable to search large domains due largely to expensive optimization techniques. In this thesis, we develop scalable sensor placement methods based on data-driven modal decompositions, and combine efficient techniques from model reduction and matrix algebra that allow us to bypass the exhaustive (brute-force or convex) optimization. This chapter first motivates the need for a data-driven approach, then provides a brief overview of existing placement methods, and finally, outlines the organization and contributions of this thesis.

1.1 Motivation

Many natural systems are complex, and existing analytic models are often insufficient for discerning regions in which to place sensors. For example, the governing equations of fluid dynamics, the nonlinear Navier-Stokes equations, characterize too broad a range of dynamical behavior – oceans, turbulence, shear flow and more – for us to discern optimal sensor and actuator locations based on the equation itself. In fields where analytic governing models are lacking such as neuroscience, the macroscale connectivity of neurons is largely studied using medical imaging to discern function, diagnose disorders, and ultimately, determine optimal sites for neural stimulation and measurement. Observation and simulation are the only tractable means to study such systems, correct assumptions in our model and derive optimal observers and controllers. Thus, a data-driven approach is proposed, that leverages global data from numerical simulations and imaging to determine where to place observers and controllers in operating conditions.

Recent years have seen an explosion of high-dimensional data generated by simulation, experiment, imaging, and even social media. Discretization in numerical simulations is growing in resolution proportional to computational power, limited only by the size of matrices that can be manipulated in memory. At the same time, imaging techniques collect data at scales that surpass data storage and transmission capabilities. The major benefit here is the fine resolution of big data resolves all scales of dynamics. However, sensor (observer) and actuator (controller) design lags behind the rate of data collection, and there are severe limitations on the number and function of sensors deployed into vehicular, robotic, and medical applications. In applications where individual sensors are expensive or have limited capabilities, minimizing the number of sensors through optimal design is critically enabling.

However, as the size of the data to be studied grows, so do the candidate search spaces for sensors. Indeed, the search complexity scales combinatorially as there are $\binom{n}{p}$ possible choices of p point sensors out of an n-dimensional state. This combinatorial growth in complexity well outpaces exponentially increasing computational power predicted by Moore's Law. Existing heuristics are increasingly unable to keep up with searching over entire domain discretizations, instead heavily downsampling the domain to smaller sets of candidate sensor locations. In doing so, these methods miss important localized dynamics and are unable to properly leverage the data available. In this thesis, we develop scalable, efficient, data-driven sensor placement algorithms that leverage compressive dimension reduction and matrix factorization techniques to keep up with the current pace of data acquisition.

1.2 Historical perspectives

To provide context for our approach, we briefly overview existing approaches for optimal sensor placement. The typical objectives that are minimized or maximized for a given sensor configuration quantify some probabilistic measure of its uncertainty, such as entropy, mutual information, or error covariance volume. It is important to note that all these criteria need to be optimized over an NP hard search of the candidate sensors, and existing approaches all develop different greedy heuristics to tackle the exhaustive search. In small search spaces it is possible to use so-called *branch-and-bound* methods to speed up the exhaustive search [164, 91], but these can handle searches of no more than tens of candidate locations.

A common data-driven alternative fits multivariate Gaussian distributions or processes to the data, then takes greedy steps to maximize different objectives, such as the mutual information or entropy characterizing a certain placement [40, 97, 131]. A more general formalism of sensor design criteria exists for parameter estimation based on minimizing certain scalarizations of the estimation errors, also called *optimal experimental design* [104, 89, 18, 123]. An important benefit of these is they are not limited to Gaussian distributions, and have been formulated for both maximum likelihood [18] and Bayesian inverse problems [84, 3]. There also exist several combined sensor and actuator placement methods designed for the various incarnations of optimal control. These attempt to optimize different properties based on linear time-invariant systems, such as observability [4, 79], controllability [142, 14], H_2 norm [43], and H_∞ norms [119]. A common solution method is to relax the subset selection problem to a convex optimization over fractional sensor weights [18, 84, 44]. Importantly, the above methods cannot computationally handle large-scale problems without subsampling the domain with grids of candidate sensors. In doing so, the underlying models miss important scales of spatiotemporal dynamics. However, despite the challenges of sensing and actuation in high-dimensional, possibly nonlinear dynamical systems, there are promising indicators that this problem may be tractable with modern techniques.

1.3 Sparsity and data-driven learning

Fortunately, we can leverage recent advances from data-driven learning to compactly represent the ambient high-dimensional space, and then determine sparse sensor locations based on the compact representation. High-dimensional systems, such as those found in fluids, epidemiology, neuroscience and power grids, typically exhibit dominant coherent structures that evolve on a low-dimensional attractor. Indeed, much of the success of modern machine learning rests on the ability to identify and take advantage of patterns and features in high-dimensional data. These low-dimensional patterns are often identified using dimensionality reduction techniques such as the proper orthogonal decomposition (POD) [13], or more recently via dynamic mode decomposition [133, 137], and diffusion maps [47]. In control theory, balanced truncation [113], balanced POD (BPOD) [134, 166] and the eigensystem realization algorithm [86] can obtain control-oriented, reduced-order models (ROM) for systems with actuation.

In addition to advances in dimensionality reduction, key developments in signal recovery capitalize on sparsity – the idea that natural signals possess only a few active components within a generic representation basis. For example, compressed sensing [58, 34] exploits the geometry of sparse vectors in high-dimensional spaces, using random measurements of a signal to recover its sparsest representation in a generic Fourier or wavelet basis, and hence reconstruct the high-dimensional signal. Compressed sensing, through its regularizing sparsity constraint, bypasses the Shannon-Nyquist sampling requirement that a signal be measured at a rate sufficient to resolve its fastest frequency [121]. Indeed, the measurement/compression requirement scales with a signal's sparsity level in the basis [156](i.e., number of nonzero Fourier coefficients). However, this requirement can be further reduced by tailoring a specific low-rank embedding to the expected signal (human faces, shear flow, etc.), and recovering its representation in the low-rank basis. In these cases it has been observed empirically that the number of measurements need only scale with the intrinsic rank of the low-rank embedding [21, 29]. Fortunately, structured signals possess a much lower intrinsic rank in the afore-mentioned data-driven bases compared to their representation in generic dictionaries. Importantly, this allows us to optimize measurement locations (sensors) based on the known low-rank structure, and minimize the number of sensors required for global inference.

Low-rank embeddings such as POD have already been used in reduced order modeling to select measurements in the state space that are informative for feature space reconstruction. These so-called *empirical interpolation methods* (EIMs) seek the best interpolation points for a given basis of POD features [10, 41, 60]. These methods have primarily been used to speed up the evaluation of nonlinear terms in high-dimensional, parametrized systems. Fortunately, the resulting interpolation points correspond to measurements in the state space, but their use for data-driven sensor placement has largely been overlooked. We will focus on this latter formulation for sensor selection and explore convex and greedy optimization methods for solving it.

1.4 Organization and contributions

This thesis explores optimized sensor placement based on tailored low-rank libraries of features extracted from training data. Optimal point sensors are discovered using the singular value decomposition and matrix QR pivoting, which are two ubiquitous numerical computations that underpin modern linear dimensionality reduction. An overview of our proposed approach, which determines optimal sensor placements efficiently by leveraging dimensionality reduction of data, is given in Figure 1.1. We provide a background on compressed sensing signal recovery and reduced-order approximation from empirical interpolation points in Chapter 2. These empirical interpolation methods provide the foundation for our sensor placement method, as the interpolation points correspond to measurement locations in a data-driven low-rank POD basis. Chapter 3 presents our pivoting based sensor placement method, which builds upon empirical interpolation methods and is formulated to minimize reconstruction error covariance. QR pivoting in particular is an extremely efficient sensor placement algorithm that has yet to be adopted by practitioners of experimental design,

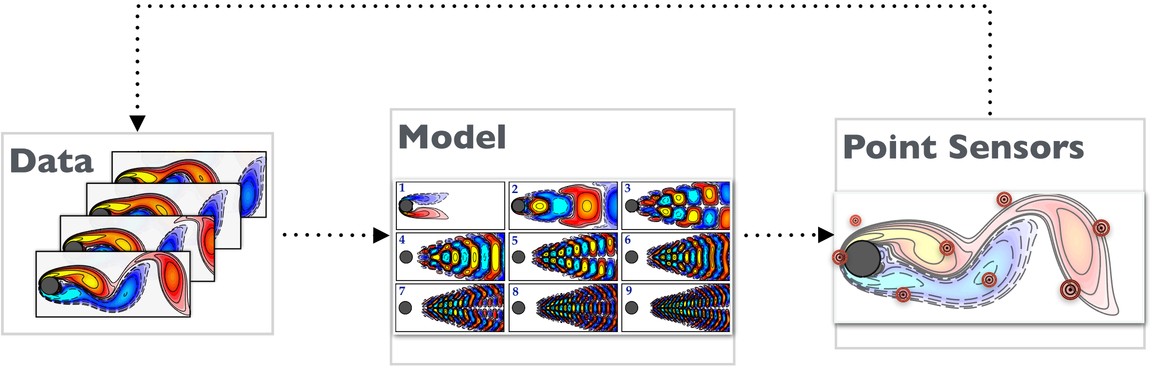


Figure 1.1: Overview of data-driven sensor placement. We exploit dimensionality reduction of high-dimensional data to efficiently determine optimized sensor placement for signal reconstruction, estimation and control.

data assimilation and optimal control. Indeed, it exceeds competing heuristic sensor selection methods by orders of magnitude in computational performance because it is formulated for efficient matrix computation. Further, analytic bounds exist for the distance to the optimum, which are typically unavailable for competing methods such as convex relaxation of sensor weights – these methods employ a nonlinear rounding procedure after computation rendering theoretical analysis difficult. We further demonstrate that signal reconstruction with the QR pivot sensors yields marginally better accuracy than the other methods surveyed, including empirical interpolation and D-optimal experimental design. This success of our method is demonstrated on actual images and satellite measurements of high-dimensional, stochastic, and possibly nonlinear systems. Even these demonstrate remarkably low-rank structure when compressing using the appropriate representation (POD, DMD or RPCA). These examples of signal reconstruction surveyed in chapters 3 and 4 have immediate impact for real-world applications in image processing, data assimilation for oceanography and climate, as well as drastically reduced assisted part scanning in manufacturing.

Identification and classification tasks are implemented using sparse approximation within larger libraries of dynamics, based on the assumption that any single regime activates only a subset of library coefficients. In chapters 5 and 6, we further adapt QR sampling to detect and classify dynamically relevant low-rank phenomena, which are contrasted to the energetic features revealed by POD. Chapter 5 demonstrated a multiresolution analysis framework using an augmented, multiresolution DMD library to optimally sample rare and intermittent dynamics. This was shown to be effective at quantifying the El Niño warming phenomenon, which has so far evaded data-driven identification using SVD-based methods. Such low-energy dynamics are either mixed or missed within POD modes. However, the DMD-based QR pivots are able to identify the El Niño region, which is an important development for measurement-based dynamical regime control in fluid dynamics, oceanography, aerodynamics and biology. These low-energy dynamical transitions are often observed in nature, and remarkably insects and birds are able to harness sparse observations of changing flow fields in flight response. Chapter 6 develops a similar framework to classify dynamical regimes in flight, based on low-rank frequency decompositions from simulated strain signals.

This work concludes with a natural extension to combined sensor and actuator placement for linear optimal control in chapter 7. QR pivoting naturally extends to this setting, in which balanced reduced order models already provide low-rank representations of system observability and controllability. Hence, actuators are computed from adjoint balanced modes, quite analogously to how sensors are computed from direct POD modes of a system. This is relevant for the high-dimensional systems we consider, for which it is advantageous to identify key sensor and actuator locations for fast, low-latency control decisions.

1.5 Conventions and Notations

This thesis borrows concepts from signal processing, fluid dynamics, control theory and machine learning, and as such there is significant overlap between standard notation. We have made every attempt to maintain consistency between notation while respecting the standard notation of each field. When there is conflict, we default to standard dynamical systems or control theory notation for the system, actuation and observation matrices. A table of standard notation used throughout this thesis is given in Table 1.1.

Nomenclature	
Scalars	
n	State dimension
m	Number of snapshots in time
p	Number of sensors (measurements)
q	Number of actuators (inputs)
r	Intrinsic numerical rank
K	Sparsity of state in universal basis $\mathbf{\Psi}$
η	Variance of zero-mean sensor noise
Vectors	
$\mathbf{x} \in \mathbb{R}^n$	High-dimensional state
$\mathbf{y} \in \mathbb{R}^p$	Measurements of state
$\mathbf{a} \in \mathbb{R}^r$	Tailored basis coefficients
$\mathbf{e}_j \in \mathbb{R}^n$	Canonical basis vectors for $\mathbb{R}^n$
$\mathbf{s} \in \mathbb{R}^n$	K-sparse basis coefficients
$\boldsymbol{\gamma} \in \mathbb{N}^p$	Sensor placement indices
$\boldsymbol{\psi} \in \mathbb{R}^n$	POD modes (columns of $\mathbf{\Psi}$)
$\boldsymbol{\phi} \in \mathbb{R}^n$	DMD modes (columns of $\mathbf{\Phi}$)
$\boldsymbol{\theta} \in \mathbb{R}^{1\times r}$	Rows of $\mathbf{\Theta}$
Matrices	
$\mathbf{B} \in \mathbb{R}^{n\times q}$	Actuation matrix
$\mathbf{C} \in \mathbb{R}^{p\times n}$	Measurement matrix
$\mathbf{Q}$	Unitary matrix in QR factorization
$\mathbf{R}$	Upper triangular matrix in QR factorization
$\mathbf{\Psi} \in \mathbb{R}^{n\times r}$	POD tailored basis
$\mathbf{\Phi} \in \mathbb{R}^{n\times r}$	DMD or adjoint modes
$\mathbf{\Theta} = \mathbf{C}\mathbf{\Psi}$	Product of measurement and basis
$\mathbf{X} \in \mathbb{R}^{n\times m}$	Data matrix with m snapshots

Table 1.1: This table presents the standard notation used throughout this thesis, unless otherwise specified.

1.6 Published work

The work presented in this thesis has resulted in the following archival and peer-reviewed publications (see below). The majority of textual and visual content in Chapters 2,3,4, and 5 are reproduced with permission from the following:

> K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems*, 38(3):63–86, June 2018. DOI: `10.1109/MCS.2018.2810460`
>
> K. Manohar, S. L. Brunton, and J. N. Kutz. Environment identification in flight using sparse approximation of wing strain. *Journal of Fluids and Structures*, 70:162–180, 2017. DOI: `10.1016/j.jfluidstructs.2017.01.008`.
>
> K. Manohar, T. Hogan, J. Buttrick, A. G. Banerjee, J. N. Kutz, and S. L. Brunton. Predicting shim gaps in aircraft assembly with machine learning and sparse sensing. *Journal of Manufacturing Systems*, 2018. DOI: `10.1016/j.jmsy.2018.01.011`.
>
> K. Manohar, E. Kaiser, S. L. Brunton, and J. N. Kutz. Optimized sampling for multiscale dynamics. *arXiv preprint*, 2018. `arXiv:1712.05085`.

Chapter 2

SPARSITY AND DIMENSIONALITY REDUCTION

There are promising indicators that the global inference of complex dynamical systems is possible from local measurements. In addition to advances in dimensionality reduction, key developments in optimization and sparse representations in high-dimensional spaces are providing powerful new techniques to obtain approximate solutions to NP-hard, combinatorially difficult problems with scalable convex optimization architectures. Key to this framework is the assumption of sparsity – the idea that natural signals possess a sparse representation in some basis. There are two ways that local measurements can exploit this property for global signal reconstruction. Sparse representations have favorable geometric properties that permit the recovery of high-dimensional signals from compressed measurements using convex optimization, a procedure known as *compressed sensing*. In our approach, an alternative construct permits a significant reduction in the number of measurements by exploiting low-rank decompositions of training data. Unlike randomized compressed sensing, optimal sensor placement fits naturally in the latter approach, where sensors can be formulated as empirical interpolation points within these low-rank, data-driven *dictionaries*. This chapter sets up background material necessary for both perspectives: 1) compressed sensing in a generic Fourier dictionary, and 2) empirical interpolation in a tailored dictionary of proper orthogonal decomposition (POD) modes. We summarize how both modalities enable global reconstruction from local measurements, and further develop the rationale behind data-driven optimal sensor placement using a tailored dictionary.

2.1 Compression in a generic dictionary

The majority of natural signals, such as images and audio, are highly compressible, meaning that when the signal is written in an appropriate coordinate system, only a few basis modes are active. These few values corresponding to the large mode amplitudes must be stored for accurate reconstruction, providing a significant reduction compared to the original signal size. In other words, in the universal transform basis, the signal may be approximated by a *sparse* vector containing mostly zeros. This inherent sparsity of natural signals is central to the mathematical framework of compressed sensing.

The theory of compressed sensing [58, 36, 37, 35, 34, 38, 8, 9] inverts this compression paradigm. Instead of collecting high-dimensional measurements just to compress and discard most of the information, it may be possible to collect a low-dimensional subsample or compression of the data and then infer the sparse vector of coefficients in the transformed coordinate system.

2.1.1 Theory of compressed sensing

Mathematically, a compressible signal $\mathbf{x} \in \mathbb{R}^n$ may be written as a sparse vector $\mathbf{s} \in \mathbb{R}^n$ in a new basis given by $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$:

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s}. \tag{2.1}$$

The vector $\mathbf{s}$ is called K-sparse if there are exactly K nonzero elements. To be able to represent *any* natural signal, rather than just those from a tailored category, the basis $\mathbf{\Psi}$ must be complete.

Consider a set of measurements $\mathbf{y} \in \mathbb{R}^p$, obtained via a measurement matrix $\mathbf{C} \in \mathbb{R}^{p \times n}$:

$$\mathbf{y} = \mathbf{C}\mathbf{x} = \mathbf{C}\mathbf{\Psi}\mathbf{s} = \mathbf{\Theta}\mathbf{s}. \tag{2.2}$$

In general, for $p < n$ (2.2) is underdetermined, and there are infinitely many solutions. The least-squares (minimum $\|\mathbf{s}\|_2$) solution is not sparse, and typically yields poor reconstruction.

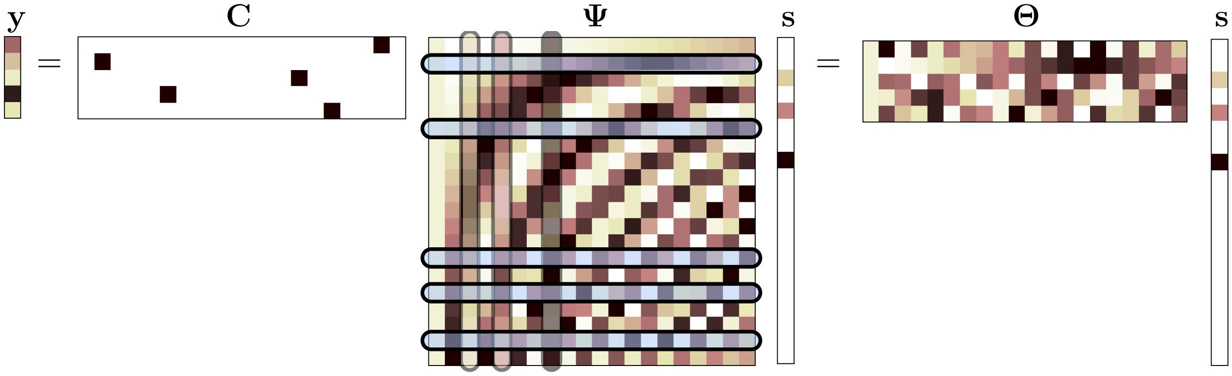


Figure 2.1: Compressed sensing provides the sparsest solution to an underdetermined linear system.

Instead, knowing that natural signals are sparse, we seek the sparsest $\mathbf{s}$ consistent with the measurements $\mathbf{y}$:

$$\mathbf{s} = \underset{\mathbf{s}'}{\arg\min} \|\mathbf{s}'\|_0, \text{ such that } \mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s}', \tag{2.3}$$

where $\|\mathbf{s}\|_0$ is the ℓ_0 pseudo-norm corresponding to the number of non-zero entries of $\mathbf{s}$. Unfortunately, this optimization problem is intractable, requiring a combinatorial brute-force search across all sparse vectors $\mathbf{s}$. A major innovation of compressed sensing is a set of conditions on the measurement matrix $\mathbf{C}$ that allow the non-convex ℓ_0-minimization in (2.3) to be relaxed to a convex ℓ_1-minimization:

$$\mathbf{s} = \underset{\mathbf{s}'}{\arg\min} \|\mathbf{s}'\|_1, \text{ such that } \mathbf{y} = \mathbf{C}\boldsymbol{\Psi}\mathbf{s}', \tag{2.4}$$

where $\|\mathbf{s}\|_1 = \sum_{k=1}^{n} |s_k|$. This formulation is shown schematically in Fig. 2.1.

For the ℓ_1-minimization in (2.4) to yield the sparsest solution in (2.3) with high probability, the measurements $\mathbf{C}$ must be chosen so that $\boldsymbol{\Theta} = \mathbf{C}\boldsymbol{\Psi}$ satisfies a *restricted isometry property* (RIP):

$$(1 - \delta_K)\|\mathbf{s}\|_2^2 \leq \|\mathbf{C}\boldsymbol{\Psi}\mathbf{s}\|_2^2 \leq (1 + \delta_K)\|\mathbf{s}\|_2^2, \tag{2.5}$$

where δ_K is a small positive *restricted isometry* constant. In particular, there are two conditions on $\mathbf{C}$ for a RIP to be satisfied for all K-sparse vectors $\mathbf{s}$:

1. The measurements $\mathbf{C}$ must be *incoherent* with respect to the basis $\boldsymbol{\Psi}$. This incoherence means that the rows of $\mathbf{C}$ are sufficiently uncorrelated with the columns of $\boldsymbol{\Psi}$, as quantified by μ:
$$\mu(\mathbf{C}, \boldsymbol{\Psi}) = \sqrt{n} \max_{j,k} \| \langle \mathbf{c}_k, \boldsymbol{\psi}_j \rangle \|. \tag{2.6}$$
Small μ indicates better incoherent measurements, with an optimal value of $\mu = 1$. Here, $\mathbf{c}_k$ denotes the k-th row of $\mathbf{C}$ and $\boldsymbol{\psi}_j$ the j-th column of $\boldsymbol{\Psi}$, both of which are assumed to be normalized.

2. The number of measurements p must satisfy:
$$p \sim \mathcal{O}(K \log(n/K)). \tag{2.7}$$
The $K \log(n/K)$ term above is generally multiplied by a small constant multiple of the incoherence. Thus, fewer measurements are required if they are less coherent.

Intuitively, the existence of a RIP implies that the geometry of sparse vectors is preserved through the measurement matrix $\mathbf{C}\boldsymbol{\Psi}$. Determining the exact constant δ_K may be extremely challenging in practice, and it tends to be more desirable to characterize the statistical properties of δ_K, as the measurement matrix $\mathbf{C}$ may be randomly chosen. **??** describes why it is not possible to use QR pivot locations as optimized sensors for compressed sensing, since they fail to identify the sparse structure of an unknown signal.

Typically, a generic basis such as Fourier or wavelets may be used to represent the signal sparsely. Spatially localized measurements (i.e., single pixels in the case of an image) are optimally incoherent with respect to the Fourier basis, so that $\mu(\mathbf{C}, \boldsymbol{\Psi}) = 1$. Thus, single pixel measurements are ideal because they excite a broadband frequency response. In contrast, a measurement corresponding to a fixed Fourier mode would be uninformative; if the signal is

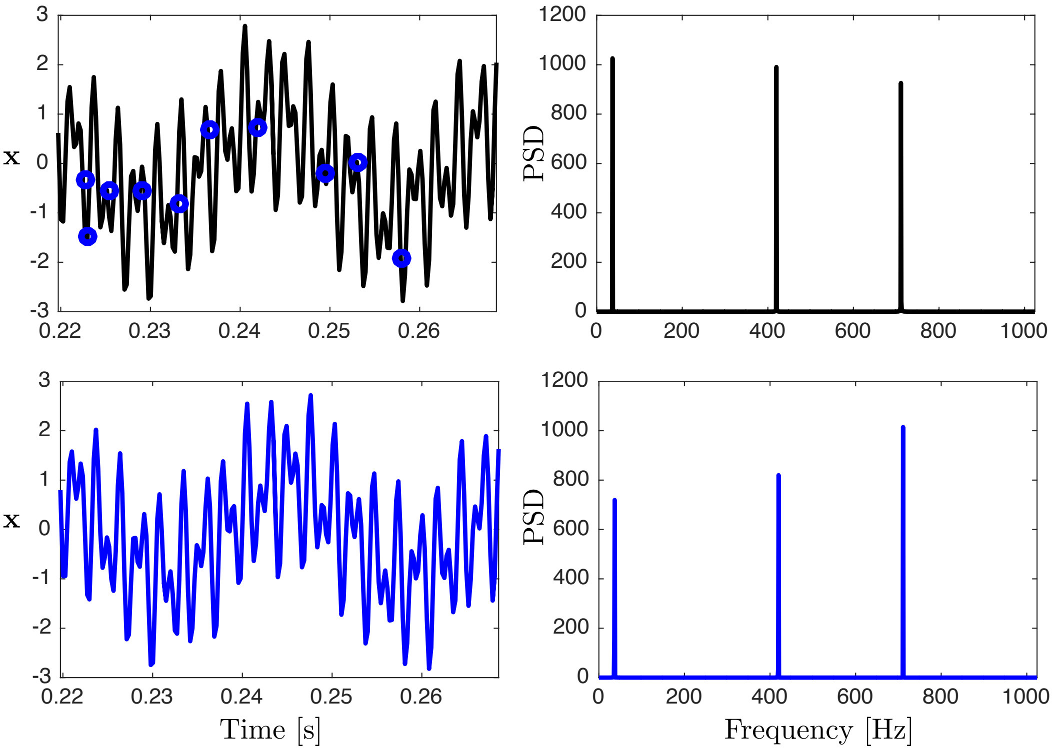


Figure 2.2: Compressed sensing, applied to three-tone signal.

not sparse in this particular frequency, this measurement provides no information about the other Fourier modes. For many engineering applications, spatially localized measurements are desirable, as they correspond to physically realizable sensors, such as buoys placed in the ocean.

One of the major results of compressed sensing is that random projection measurements of the state (i.e., entries of $\mathbf{C}$ that are Bernoulli or Gaussian random variables) are incoherent with respect to nearly any generic basis $\boldsymbol{\Psi}$. This result is truly remarkable; however, the incoherence of random projections is not optimal, and typically scales as $\mu \sim \sqrt{2\log(n)}$. Moreover, it may be difficult to obtain random projections of the full state $\mathbf{x}$ in physical applications.

There are many alternative strategies to solve for the sparsest solution to (2.2). Greedy algorithms are often used [150, 151, 153, 155, 152, 154, 156, 70, 71], including the compressed sampling matching pursuit (CoSaMP) algorithm [118].

2.2 Interpolation in a tailored dictionary

Compressed sensing requires measurements that number on the scale of the signal's sparsity level in a universal dictionary. Because generic dictionaries are designed to represent a broad class of signals (they are *complete*), more measurements are needed than if we had basis functions tailored to the expected data. Data arising from physical systems typically possess extremely low-rank spatiotemporal correlation structure that can be exploited to drastically reduce the number of sensors required for global inference. This minimal requirement is equivalent to the intrinsic rank of the data, which is much smaller than its rank in a universal dictionary such as Fourier modes. In this section we describe how to construct this optimal low-rank representation using a dictionary of proper orthogonal decomposition (POD) modes. Finally, we introduce empirical interpolation methods that locate the states that are maximally informative of the underlying physical structure, which can then be used for strategic sensor placement.

2.2.1 Proper Orthogonal Decomposition

Consider data generated from a complex dynamical system formulated in discrete time

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k), \tag{2.8}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a high-dimensional vector of state variables. The large number of states typically results from numerical discretization or high-resolution data acquisition methods, i.e., satellite, particle-image velocimetry, etc. Assume we can access a data matrix of m *snapshots* of the system in time

$$\mathbf{X} = \left[\mathbf{x}_1 \mid \mathbf{x}_2 \mid \ \ldots \mathbf{x}_m\right]. \tag{2.9}$$

The proper orthogonal decomposition (POD) [13] can be applied to compute an optimal low-rank approximation to the data. The POD constructs a set of orthonormal vectors $\{\boldsymbol{\psi}_k\}_{k=1}^r$ that minimizes the following approximation objective

$$\text{minimize}_{\{\boldsymbol{\psi}_k\}} \sum_{k=1}^{m} \|\mathbf{x}_k - (\mathbf{x}_k^T \boldsymbol{\psi}_k)\boldsymbol{\psi}_k\|_2. \tag{2.10}$$

Note that the objective being minimized is equivalent to the Frobenius norm error between the data matrix and its optimal rank r approximant

$$\sum_{k=1}^{m} \|\mathbf{x}_k - (\mathbf{x}_k^T \boldsymbol{\psi}_k)\boldsymbol{\psi}_k\|_2 = \|\mathbf{X} - \boldsymbol{\Psi}_r \boldsymbol{\Psi}_r^T \mathbf{X}\|_F, \tag{2.11}$$

where $\boldsymbol{\Psi}_r = [\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_r]$. According to the Eckart-Young-Mirsky Theorem [62], the solution is given by the singular value decomposition. Specifically, the projector $\boldsymbol{\Psi}_r$ is given by the first r left singular vectors of $\mathbf{X}$. The POD modes (columns of $\boldsymbol{\Psi}_r$) are eigenvectors of the data covariance matrix, and are therefore equivalent to principal component analysis or PCA [125], the Karhunen-Loève expansion [90], and empirical orthogonal functions [105]. The only difference is that POD is formulated explicitly for time-dependent dynamical systems.

2.2.2 Empirical Interpolation Methods

POD constructs a low-rank orthogonal basis for a subspace in Hilbert space. A common model reduction method known as *Galerkin projection* is to project dynamics into this subspace using POD modes to obtain a *reduced order model* or ROM. Consider the following model for a high-dimensional dynamical system ($\mathbf{x} \in \mathbb{R}^n$)

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(\mathbf{x}), \tag{2.12}$$

in which nonlinear dynamics $\mathbf{f}(\mathbf{x})$ are disambiguated from the linear dynamics. Given sufficiently many snapshots of the high-dimensional states, we can compute the POD modes and hence a Galerkin projection to obtain a low-dimensional model of the system evolution

$$\frac{d\mathbf{a}(t)}{dt} = (\mathbf{\Psi}^*\mathbf{A}\mathbf{\Psi})\mathbf{a}(t) + \mathbf{\Psi}^*\mathbf{f}(\mathbf{x}(t)), \tag{2.13}$$

where $\mathbf{a} = \mathbf{\Psi}^*\mathbf{x}$ is the projected r-dimensional state in the low-dimensional subspace. Typically $r \ll n$, which drastically reduces the computational resources required for time-stepping the r-dimensional reduced system instead of the full order model. However, nonlinear terms are still evaluated in the high-dimensional space before being projected, and these evaluations become computationally prohibitive for high-dimensional full order models.

Recent methods bypass the evaluation of high-dimensional nonlinearities by interpolating them at special optimal interpolation points computed from POD basis functions. These points are analogous to optimal Fekete and Chebyshev points for polynomial interpolation, with the exception of being dependent on the data. As we shall see, these points effectively correspond to data-driven optimal sensor locations, and hence the interpolation is known as *empirical* interpolation method or EIM [10, 41]. EIM computes r optimal points (states) for interpolation in a basis of POD modes that span the nonlinear terms. That is, given r POD modes $\mathbf{\Psi}_{NL}$ that span the nonlinear evolution $\mathbf{X} = [\mathbf{f}(\mathbf{x}(t_1)) \ldots \mathbf{f}(\mathbf{x}(t_m))]$, EIM computes the optimal r indices at which to interpolate the nonlinearity $\mathbf{f}(\mathbf{x})$. Given r interpolating basis functions $\{\boldsymbol{\psi}_k\}_{k=1}^r$ and the vector of the nonlinear values at the interpolating points $\mathbf{Cf}$, then the interpolant evaluates to

$$\hat{\mathbf{f}} = \mathbf{\Psi}_{NL}(\mathbf{C}\mathbf{\Psi}_{NL})^{-1}(\mathbf{Cf}). \tag{2.14}$$

This equation is entirely analogous to polynomial interpolation if the basis $\mathbf{\Psi}_r$ is the Vandermonde matrix of monomials on $\mathbf{x}$. The interpolation point sampling operator is denoted $\mathbf{C}$ to be consistent with notation in subsequent chapters. EIM constructs the sampling operator to minimize the approximation error between the nonlinearity and its interpolant

$$\text{minimize}_{\mathbf{C}} \|\mathbf{f} - \mathbf{\Psi}_{NL}(\mathbf{C}\mathbf{\Psi}_{NL})^{-1}\mathbf{Cf}\|_2. \tag{2.15}$$

Algorithm 1 Discrete empirical interpolation method [41]
Input: POD modes $\mathbf{\Psi} \in \mathbb{R}^{n\times r}$
Output: Interpolation points $\boldsymbol{\gamma}, \mathbf{C}$

```
procedure DEIM( $\mathbf{\Psi}$ )
    $\gamma_1 \leftarrow \text{argmax}_{j=1,\dots,n}(\boldsymbol{\psi}_1)_j$
    $\mathbf{C} = [\mathbf{e}_{\gamma_1}^T]$
    $\hat{\mathbf{\Psi}} = [\boldsymbol{\psi}_1]$
    for $k = 2, \dots, r$ do
        Solve $(\mathbf{C}\hat{\mathbf{\Psi}})\mathbf{c} = \mathbf{C}\boldsymbol{\psi}_k$ for $\mathbf{c}$
        $\mathbf{v} = \boldsymbol{\psi}_k - \hat{\mathbf{\Psi}}\mathbf{c}$
        $\gamma_k = \text{argmax}_{j \notin \boldsymbol{\gamma}} v_j$
        $\boldsymbol{\gamma} \leftarrow [\boldsymbol{\gamma},\ \gamma_k]$, $\mathbf{C} = [\mathbf{C}^T, \mathbf{e}_{\gamma_k}]^T$
        $\hat{\mathbf{\Psi}} = [\hat{\mathbf{\Psi}}, \boldsymbol{\psi}_k]$
return $\boldsymbol{\gamma}$, $\mathbf{C}$
```

Algorithm 1 summarizes the greedy computational method for computing discrete EIM or DEIM points. After initializing the first interpolation point at the maximum of the first POD mode, DEIM inductively attempts to minimize the approximation error between the $k-1$ mode interpolant and the k^{th} POD mode (which is orthonormal to the others). In this manner the interpolation points become maximally informative of the POD basis. DEIM greedily selects indices $\boldsymbol{\gamma}$ to minimize growth of the spectral norm of $\|(\mathbf{C}\mathbf{\Psi})^{-1}\|_2$. In the original DEIM paper [41], Chaturantabut and Sorensen show that the projection error is controlled by this term in the following way

Lemma 1 (Chaturantabut and Sorensen [41])**.** *Given orthonormal* $\mathbf{\Psi} \in \mathbb{R}^{n\times r}$ *(*$r \ll n$*) and the DEIM projection of an arbitrary* $\mathbf{f} \in \mathbb{R}^n$ *defined by*

$$\hat{\mathbf{f}} \triangleq \mathbf{\Psi}(\mathbf{C}\mathbf{\Psi})^{-1} \cdot \mathbf{C}\mathbf{f}, \tag{2.16}$$

with $\mathbf{C}$ *computed by algorithm 1, the interpolation error satisfies the following upper bound*

$$\|\mathbf{f} - \hat{\mathbf{f}}\|_2 \leq \|(\mathbf{C}\mathbf{\Psi})^{-1}\|_2 \cdot \|(\mathbb{I} - \mathbf{\Psi}\mathbf{\Psi}^*)\mathbf{f}\|_2, \tag{2.17}$$

where

$$\|(\mathbf{C}\mathbf{\Psi})^{-1}\|_2 \leq \frac{(1+\sqrt{2n})^{r-1}}{\|\boldsymbol{\psi}_1\|_\infty}. \tag{2.18}$$

Because the second term ($\|(\mathbb{I} - \mathbf{\Psi}\mathbf{\Psi}^*)\mathbf{f}\|_2$) is already the minimal approximation error for $\mathbf{f}$ in the range of $\mathbf{\Psi}$, we can focus on the term that depends on the measurements (interpolation points $\mathbf{C}$). Minimizing the measurement-dependent term is the goal of empirical interpolation. The next chapter presents an improvement of DEIM that possesses a tighter upper bound than eq. (2.18). We show in subsequent chapters how these EIMs can be applied more broadly for data-driven sensor placement of any system, static or time dependent.

Chapter 3

SENSOR PLACEMENT FOR SIGNAL RECONSTRUCTION

Compressed sensing is ideal for the recovery of a high-dimensional signal of *unknown* content using random measurements in a universal basis. However, if information is available about the *type* of signal (such as the signal is a turbulent velocity field or an image of a human face), it is possible to design optimized sensors that are tailored for the particular signals of interest. We have seen previously that n-dimensional states can be reconstructed faithfully with $r \ll n$ measurements, which is no more than the intrinsic rank of data in POD feature space. These low-rank features, mined from patterns in the data, facilitate the design of specialized sensors that are tailored to a specific problem. Crucial to our approach is the assumption that high-dimensional states can be expressed as a linear combination of basis features using a low-dimensional vector $\mathbf{a} \in \mathbb{R}^r$ of basis coefficients

$$\mathbf{x} = \sum_{k=1}^{r} a_k \boldsymbol{\psi}_k = \boldsymbol{\Psi}_r \mathbf{a}, \tag{3.1}$$

where $\mathbf{a}$ is a r-dimensional vector. Assuming we are only given r POD modes $\boldsymbol{\Psi}_r$ and sparse point observations of $\mathbf{x}$, we can recover $\mathbf{x}$ from its estimated POD coefficients $\mathbf{a}$.

Although this concept is closely related to the optimization of empirical interpolation points, one major shortcoming of DEIM interpolation is that it cannot be generalized to regression. That is, DEIM does not allow the selection of $p > r$ interpolation points in a basis of r features. This is problematic because parameter estimation problems become numerically more well conditioned with more equations (observations) than unknowns. Although more principled sensor selection algorithms based on POD interpolation exist, they are computationally much more expensive since they require greedy searches over the domain

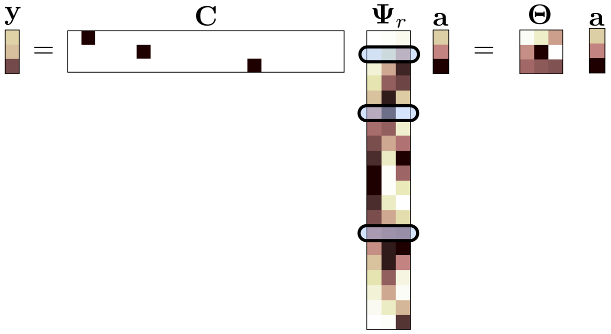


Figure 3.1: Full state reconstruction of $\mathbf{x}$ from point observations ($\mathbf{y}$) is accomplished using least squares estimation of POD coefficients ($\mathbf{a} = (\mathbf{C}\boldsymbol{\Psi}_r)^\dagger \mathbf{y}$).

to optimize the condition number.

In this chapter, we introduce gappy POD, which is a more general framework for signal reconstruction than empirical POD interpolation. Based on this, we frame an optimal sensor placement objective which is closely related to methods in Bayesian optimal experiment design. A new sensor placement method is proposed using QR pivots of POD modes, which has been previously formulated for empirical interpolation. We generalize this method to select more sensors than POD modes (when there are more observations than unknowns), and the method is shown to solve the D-optimal experimental design criterion. Finally, we compute sensors and reconstruct a variety of high-dimensional static and dynamic datasets including the Yale B facial image dataset and weekly sea surface temperature fields.

3.1 Gappy POD reconstruction

Gappy POD is a signal recovery formulation using POD basis functions and incomplete measurements. Originally defined for facial image recovery using POD modes of faces (so-called *eigenfaces*), gappy POD has gained popularity for reconstructing flow fields from sensor measurements. Its initial introduction featured random projections of face data much like compressed sensing, however it was quickly discovered that gappy recovery yields better performance with optimized measurements.

Problem statement. Consider taking $p > r$ linear observations of an n-dimensional state vector that is low-rank in POD basis $\boldsymbol{\Psi}_r$, where each observation may be corrupted by Gaussian i.i.d. white noise $\boldsymbol{\zeta} \sim N(0, \eta^2)$,

$$\begin{aligned} \mathbf{y} &= \mathbf{C}\mathbf{x} + \boldsymbol{\zeta} && (3.2) \\ &= \mathbf{C}\boldsymbol{\Psi}_r\mathbf{a} + \boldsymbol{\zeta}, && (3.3) \end{aligned}$$

and point sensor measurements impose the following structure on the measurement matrix $\mathbf{C} \in \mathbb{R}^{p\times n}$

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{e}_{\gamma_1} & \mathbf{e}_{\gamma_2} & \dots & \mathbf{e}_{\gamma_p} \end{bmatrix}^T. \tag{3.4}$$

Here $\mathbf{e}_j$ are the canonical basis vectors for $\mathbb{R}^n$ with a unit entry at index j and zeros elsewhere. The observations consist of p elements selected from $\mathbf{x}$

$$\mathbf{C}\mathbf{x} = [x_{\gamma_1}\ x_{\gamma_2}\ \dots\ x_{\gamma_p}]^T, \tag{3.5}$$

where $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_p\} \subset \{1, \dots, n\}$ denotes the index set of sensor locations with cardinality $|\boldsymbol{\gamma}| = p$. Furthermore, the number of measurements is contrained to be greater than or equal to the number of basis functions (i.e., $p \geq r$) for a well-defined linear inverse problem. Specifically, the recovery of $\mathbf{x}$ is reduced to a much lower-dimensional problem of estimating its basis coefficients $\mathbf{a}$. A schematic of gappy recovery using POD modes $\boldsymbol{\Psi}_r$ is shown in Fig. 3.1.

The maximum likelihood estimate of $\mathbf{a}$ is also the familiar least-squares error minimizer, which can be explicitly computed using the Moore-Penrose pseudoinverse

$$\hat{\mathbf{a}} \triangleq (\mathbf{C}\boldsymbol{\Psi}_r)^\dagger\mathbf{y}.$$

Finally, the full state $\mathbf{x}$ can be recovered as a linear combination of the estimated coefficients and POD modes

$$\hat{\mathbf{x}} \triangleq \boldsymbol{\Psi}_r \hat{\mathbf{a}}, \text{ where } \hat{\mathbf{a}} \triangleq \begin{cases} (\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\mathbf{y}, & p = r, \\ (\mathbf{C}\boldsymbol{\Psi}_r)^{\dagger}\mathbf{y}, & p > r \end{cases}. \tag{3.6}$$

Upon expanding $\mathbf{y}$ and $\hat{\mathbf{a}}$, it is clear that this estimate is equivalent to the DEIM approximation when when $p = r$ and $\mathbf{C}$ are DEIM interpolation points.

3.2 Optimal sensor placement objective

As before, the mathematical objective of sensor selection is to quantify the approximation error, $\|\mathbf{a} - \hat{\mathbf{a}}\|_2$, as a function of the measurement locations, then formulate the optimal sensor selection $\mathbf{C}$ as the unique minimizer of this expected error. We quantify the probability distribution of the approximation error using the matrix of error covariance between $\mathbf{a}$ and the maximum likelihood estimate $\hat{\mathbf{a}}$

$$\boldsymbol{\Xi} = \text{Var}(\mathbf{a} - \hat{\mathbf{a}}) = \eta^2[(\mathbf{C}\boldsymbol{\Psi}_r)^T\mathbf{C}\boldsymbol{\Psi}_r]^{-1}. \tag{3.7}$$

The error covariance $\boldsymbol{\Xi}$ characterizes the minimum volume *ρ-confidence ellipsoid*, ε_ρ, that contains the least-squares estimation error, $\mathbf{a} - \hat{\mathbf{a}}$, with probability ρ. The set of optimal sensors minimizes the volume of ε_ρ

$$\text{vol}(\varepsilon_\rho) = \delta_{\rho,r} \det \boldsymbol{\Xi}^{\frac{1}{2}}, \tag{3.8}$$

where $\delta_{\rho,r}$ only depends on ρ and r, by minimizing the determinant, which is the only sensor dependent term. This optimization is equivalent to maximizing the determinant of the inverse

$$\underset{\mathbf{C}}{\text{maximize}} \quad \log\det\left[(\mathbf{C}\boldsymbol{\Phi}_r)^T\mathbf{C}\boldsymbol{\Phi}_r\right], \tag{3.9}$$

which is commonly known as the *D-optimal* criterion in design of experiments. Recall the structure of the measurement matrix $\mathbf{C} \in \mathbb{R}^{p\times n}$ is similar to the interpolating matrix from

DEIM

$$\mathbf{C} \triangleq [\mathbf{e}_{\gamma_1} \mid \mathbf{e}_{\gamma_2} \mid \ \ldots \ \mid \mathbf{e}_{\gamma_p}]^T, \tag{3.10}$$

where $\gamma_i \in \{1, \ldots, n\}$ indexes the high-dimensional measurement space, and $\mathbf{e}_{\gamma_i}$ are the canonical unit vectors consisting of all zeros except for a unit entry at γ_i. This guarantees that each row of $\mathbf{C}$ only measures from a single spatial location, corresponding to a point sensor.

The optimal sensor configuration shrinks all components of the reconstruction error via the associated ellipsoid *volume*, i.e. the choice that minimizes the determinant of $\boldsymbol{\Xi}$. This subset optimization is a combinatorial search over $\binom{n}{p}$ candidate locations which quickly becomes computationally intractable even for moderately large n and p. Fortunately, an extremely efficient greedy optimization method for (3.9) is given by the pivoted matrix QR factorization of $\boldsymbol{\Psi}_r^T$. QR pivoting has been extensively used to compute optimal quadrature and polynomial interpolation points [33, 139, 138], as well as empirical interpolation points using an improved DEIM called Q-DEIM [60]. QR pivoting is shown to outperform related methods for optimal selection in either computational accuracy or runtime, sometimes both. These suboptimal related methods include EIM, DEIM, information theoretic criteria [97] and convex optimization [84, 18]. We now describe our sensor placement method given by the matrix QR factorization with column pivoting of $\boldsymbol{\Psi}_r^T$, which builds upon the Q-DEIM method [60].

3.3 QR pivoting method

The reduced matrix QR factorization with column pivoting decomposes a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ into a unitary matrix $\mathbf{Q}$, an upper-triangular matrix $\mathbf{R}$ and a column permutation matrix $\mathbf{C}$ such that $\mathbf{A}\mathbf{C}^T = \mathbf{Q}\mathbf{R}$. Recall that the determinant of a matrix, when expressed as a product of a unitary factor and an upper-triangular factor, is the product of the diagonal

Algorithm 2 QR factorization with column pivoting of $\mathbf{A} \in \mathbb{R}^{n \times m}$
Greedy optimization for placing p sensors $\boldsymbol{\gamma}$ from multiscale modes.
Usage: QRPIVOT($\boldsymbol{\Psi}_r^T$, p) (if $p = m$)
QRPIVOT($\boldsymbol{\Psi}_r \boldsymbol{\Psi}_r^T$, p) (if $p > m$)

```
procedure QRPIVOT( $\mathbf{A}$, $p$ )
    $\boldsymbol{\gamma} \leftarrow [\ ]$
    for $k = 1, \ldots, p$ do
        $\gamma_k = \operatorname{argmax}_{j \notin \boldsymbol{\gamma}} \|\mathbf{a}_j\|_2$
        Find Householder $\tilde{\mathbf{Q}}$ such that $\tilde{\mathbf{Q}} \cdot \begin{bmatrix} a_{kk} \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} \star \\ 0 \\ \vdots \\ 0 \end{bmatrix}$    ▷ $\star$'s represent nonzero diagonal entries in $\mathbf{R}$
        $\mathbf{A} \leftarrow \operatorname{diag}(I_{k-1}, \tilde{\mathbf{Q}}) \cdot \mathbf{A}$    ▷ Remove from all columns the orthogonal projection onto $\mathbf{a}_{\gamma_k}$
        $\boldsymbol{\gamma} \leftarrow [\boldsymbol{\gamma},\ \gamma_k]$
    return $\boldsymbol{\gamma}$
```

entries in the upper-triangular factor:

$$\left|\det \mathbf{A}\mathbf{C}^T\right| = |\det \mathbf{Q}||\det \mathbf{R}| = \prod_i |r_{ii}|, \tag{3.11}$$

The pivoted QR permutes the matrix $\boldsymbol{\Psi}_r^T$ with $\mathbf{C}^T$ to enforce the following diagonal dominance structure on the diagonal entries of its $\mathbf{R}$ factor [60]:

$$\sigma_i^2 = |r_{ii}|^2 \geq \sum_{j=i}^{k} |r_{jk}|^2; \quad 1 \leq i \leq k \leq m. \tag{3.12}$$

Column pivoting iteratively increments the volume of the pivoted submatrix by selecting a new pivot column with maximal 2-norm, then subtracting from every other column its orthogonal projection onto the pivot column, and is summarized in Algorithm 2. In this manner the QR factorization with column pivoting yields p point sensor indices (pivots) that best characterize dominant dynamical modes $\mathbf{\Psi}_r$

$$\mathbf{\Psi}_r^T \mathbf{C}^T = \mathbf{QR}. \tag{3.13}$$

Drmac and Gugercin [60, Theorem 2.1] derive an upper bound for the interpolation error that is much tighter than that of DEIM (2.18):

Lemma 2 (Drmac and Gugercin [60])**.** *Given orthonormal* $\mathbf{\Psi} \in \mathbb{R}^{n \times r}$ *($r \ll n$) and the Q-DEIM projection of an arbitrary* $\mathbf{f} \in \mathbb{R}^n$*,*

$$\hat{\mathbf{f}} \triangleq \mathbf{\Psi}(\mathbf{C\Psi})^{-1} \cdot \mathbf{Cf}, \tag{3.14}$$

with $\mathbf{C}$ *computed by algorithm 2, the interpolation error satisfies the following upper bound*

$$\|\mathbf{f} - \hat{\mathbf{f}}\|_2 \leq \|(\mathbf{C\Psi})^{-1}\|_2 \cdot \|(\mathbb{I} - \mathbf{\Psi\Psi}^*)\mathbf{f}\|_2, \tag{3.15}$$

where

$$\|(\mathbf{C\Psi})^{-1}\|_2 \leq \sqrt{n - r + 1}\frac{\sqrt{4^r + 6r - 1}}{3}. \tag{3.16}$$

Extension to oversampling. The standard pivoting formulation yields only as many pivots as there are columns of $\mathbf{\Psi}_r$ (modes). However, *oversampling*, which refers to sampling more observations than modes, promotes robustness to measurement noise, thus regularizing the reconstruction problem (3.6). An original contribution of our work is the extension of this QR pivoting method from interpolation to gappy POD regression. In the following, the product of the measurement and POD basis is denoted by $\mathbf{\Theta} = \mathbf{C\Psi}_r$. The oversampled sensor placement $p > r$ can be computed from the pivoted QR factorization of $\mathbf{\Psi}_r\mathbf{\Psi}_r^T$, where the column pivots are selected from n candidate state space locations based on the observation that the singular values of the product $\mathbf{\Theta}^T\mathbf{\Theta}$ coincide with the first r singular values of its transpose and the diagonal entries of the **R** factor:

$$\det \mathbf{\Theta}^T\mathbf{\Theta} = \prod_{i=1}^{r} \sigma_i(\mathbf{\Theta}\mathbf{\Theta}^T). \tag{3.17}$$

Here we drop the absolute value since the determinant of $\mathbf{\Theta}\mathbf{\Theta}^T$ is nonnegative. Thus, this pivoting formulation automatically optimizes the desired determinant (3.9).

$$(\mathbf{\Psi}_r\mathbf{\Psi}_r^T)\mathbf{C}^T = \mathbf{QR}. \tag{3.18}$$

The oversampled case requires an expensive QR factorization of an $n \times n$ matrix, whose storage requirements scale quadratically with state dimension. However, this operation may be advantageous for several reasons. The row selection given by the first p QR pivots increase the leading r singular values of $\mathbf{\Theta}\mathbf{\Theta}^T$, hence increasing $\det \mathbf{\Theta}^T\mathbf{\Theta}$. This is the same maximization objective used in D-optimal experiment design [124], which is typically solved with Newton iterations using a convex relaxation of the subset selection objective. These methods require one matrix factorization of an $n \times n$ matrix per iteration, leading to a runtime cost per iteration of at least $O(n^3)$. The entire procedure must be recomputed for each new choice of p. Our proposed method only requires a single $O(n^3)$ QR factorization and results in a hierarchical list of all n total pivots, with the first p pivots optimized for reconstruction in $\mathbf{\Psi}_r$ for any $p > r$. Thus, additional sensors may be leveraged if available.

The QR factorization is implemented and optimized in most standard scientific computing packages and libraries, including Matlab, LAPACK, NumPy, among many others. In addition to software-enabled acceleration, QR runtime can be significantly reduced by terminating the procedure after the first p pivots are obtained. The operation can be accelerated further using randomized techniques, for instance, by the random selection of the next pivot [60] or by using random projections to select blocks of pivots [110, 109, 61].

Our sensor placement formulation is related to the machine learning concepts of variable and feature selection [149, 76]. Such sensor (feature) selection concepts generalize to data-driven classification. For image classification using linear discriminant analysis (LDA), sparse sensors may be selected that map via low-rank modes into the discriminating subspace [23]. Moreover, sparse classification within libraries of POD modes [170, 21, 29] can be improved by augmenting empirical interpolation methods with a genetic algorithm [136].

3.4 Connection to optimal experimental design

The matrix volume objective described above is closely related to D-optimal experiment design [124]; in fact, the two problems are identical when regarding the tailored basis $\boldsymbol{\Psi}_r$ as a set of n candidate experiments of a low-dimensional subspace. Classical experimental design selects the best p out of n candidate experiments for estimating r unknown parameters $\mathbf{a} \in \mathbb{R}^r$. Each experiment, denoted $\boldsymbol{\theta}_i$, produces one output y_i that may include zero-mean i.i.d. Gaussian noise $\boldsymbol{\zeta} \sim \mathcal{N}(0, \eta^2)$. Again, we wish to estimate the parameters from p experiment outputs $\mathbf{y} \in \mathbb{R}^p$ in the following linear system,

$$\mathbf{y} = \begin{bmatrix} —\boldsymbol{\theta}_1— \\ —\boldsymbol{\theta}_2— \\ \vdots \\ —\boldsymbol{\theta}_p— \end{bmatrix} \cdot \mathbf{a} + \boldsymbol{\zeta} = \sum_{k=1}^{r} \begin{bmatrix} \Psi_{\gamma_1,k} \\ \Psi_{\gamma_2,k} \\ \vdots \\ \Psi_{\gamma_p,k} \end{bmatrix} a_k + \boldsymbol{\zeta} = \mathbf{C}\boldsymbol{\Psi}_r\mathbf{a} + \boldsymbol{\zeta}, \tag{3.19}$$

which is equivalent to the state reconstruction formulation of gappy POD [65]. In Matlab notation we sometimes refer to $\mathbf{C}\boldsymbol{\Psi}_r$ as $\boldsymbol{\Psi}_r(\boldsymbol{\gamma}, :)$. Each possible experiment $\boldsymbol{\theta}_i$ may be regarded as a row of $\boldsymbol{\Theta}$ or of the tailored basis $\boldsymbol{\Psi}_r$ such that $\boldsymbol{\theta}_i = \boldsymbol{\Psi}_r(\gamma_i, :)$. Equivalently each $\boldsymbol{\theta}_i$ is a weighted "measurement" of the lower dimensional POD parameter space (not to be confused with the point measurement operation $\mathbf{C}$). Note that when all experiments are selected the output is simply the state vector $\mathbf{x}$ since $\mathbf{x} = \boldsymbol{\Psi}_r\mathbf{a}$. Given experiment selections indexed by $\boldsymbol{\gamma}$, the estimation error covariance is equivalent to the expression for $\boldsymbol{\Xi}$ as expected

$$\mathrm{Var}(\mathbf{a} - \hat{\mathbf{a}}) = \eta^2[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^{-1} = \eta^2[(\mathbf{C}\boldsymbol{\Psi}_r)^T\mathbf{C}\boldsymbol{\Psi}_r]^{-1}. \tag{3.20}$$

D-optimal subset selection minimizes the error covariance by maximizing the matrix volume of $\mathbf{\Theta}^T\mathbf{\Theta}$:

$$\begin{aligned}\boldsymbol{\gamma}_\star &= \underset{\boldsymbol{\gamma}, |\boldsymbol{\gamma}|=p}{\operatorname{argmax}} \log\det \sum_{i=1}^{p} \boldsymbol{\theta}_i^T \boldsymbol{\theta}_i \\ &= \underset{\boldsymbol{\gamma}, |\boldsymbol{\gamma}|=p}{\operatorname{argmax}} \det(\mathbf{C}\mathbf{\Psi}_r)^T \mathbf{C}\mathbf{\Psi}_r, \end{aligned} \tag{3.21}$$

which is equivalent to (3.9). While the D-optimal criterion is formulated to shrink all components of the error via its volume, A-optimal and E-optimal design criteria optimize the trace and spectral radius of $\mathbf{\Theta}^T\mathbf{\Theta}$, hence minimizing the mean squared error and worst-case error component, respectively. The exact solutions of these optimization problems are intractable, and they are generally solved using heuristics. This is most commonly accomplished by solving the convex relaxation with a linear constraint on *sensor weights* $\boldsymbol{\beta}$,

$$\begin{aligned}\boldsymbol{\beta}_\star = \underset{\boldsymbol{\beta}\in\mathbb{R}^n}{\operatorname{argmax}} \log\det \sum_{i=1}^{n} \beta_i \boldsymbol{\theta}_i^T \boldsymbol{\theta}_i, \\ \text{subject to} \sum_{\substack{i=1 \\ 0\le\beta_i\le 1}}^{n} \beta_i = p. \end{aligned} \tag{3.22}$$

The optimized sensors are obtained by selecting the largest sensor weights from $\boldsymbol{\beta}$. The iterative methods employed to solve this problem, i.e., convex optimization and semidefinite programs [18, 84], require $O(n^3)$ matrix operations in each iteration to compute intermediate Hessians. Therefore they are computationally more expensive than the QR pivoting methods, which cost one matrix factorization in total. Greedy sampling methods such as DEIM and QR are practical for sensor placement within a large number of candidate locations over high-dimensional spatial grids; hence, they are the methods of choice in reduced-order modeling [12]. The various optimization methods for data-driven sensor selection are summarized in Table 3.1.

Method	**Objective**	**Runtime**
D-optimal subset selection	$\boldsymbol{\gamma}_\star = \underset{\gamma, \lvert\gamma\rvert = p}{\operatorname{argmax}} \log\det \sum_{i=1}^{p} \boldsymbol{\theta}_i^T \boldsymbol{\theta}_i$ $= \underset{\gamma, \lvert\gamma\rvert = p}{\operatorname{argmax}} \det(\mathbf{C}\boldsymbol{\Psi}_r)^T \mathbf{C}\boldsymbol{\Psi}_r$	$\binom{n}{p}$ determinant evaluations
Convex optimization	$\boldsymbol{\beta}_\star = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \log\det \sum_{i=1}^{n} \beta_i \boldsymbol{\theta}_i^T \boldsymbol{\theta}_i$ subject to $\sum_{\substack{i=1 \\ 0 \le \beta_i \le 1}}^{n} \beta_i = p$	$\mathcal{O}(n^3)$ per iteration
QR pivoting (greedy)	Case $p = r$: $\boldsymbol{\Psi}_r^T \mathbf{C}^T = \mathbf{QR}$ Case $p > r$: $(\boldsymbol{\Psi}_r \boldsymbol{\Psi}_r^T)\mathbf{C}^T = \mathbf{QR}$	$\mathcal{O}(nr^2)$ $\mathcal{O}(n^3)$

Table 3.1: Summary of sensor placement methods

3.5 Results and discussion

Sensor selection and signal reconstruction algorithms are implemented and compared on data from fluid dynamics, facial images, and ocean surface temperatures. The examples span a wide range of complexity, exhibit both rapid and slow singular value decay, and come from both static and dynamic systems.

In each example, optimized sensors obtained in a tailored basis with QR pivots outperform random measurements in a universal basis using compressed sensing for signal reconstruction. Moreover, for the same reconstruction performance, many fewer QR sensors are required, decreasing the cost associated with purchasing, placing, and maintaining sensors, as well as reducing the latency in computations. Thus, for a well-scoped reconstruction task with sufficient training data, we advocate principled sensor selection rather than compressed sensing. When the structure of the underlying signal is unknown, then compressed sensing provides more flexibility with an associated increase in the number of sensors.

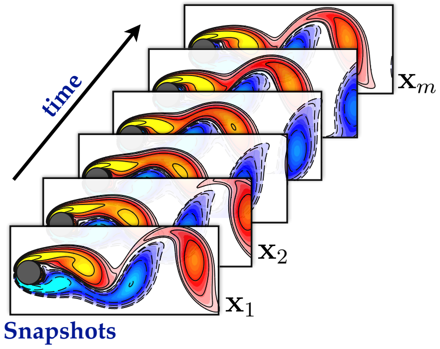
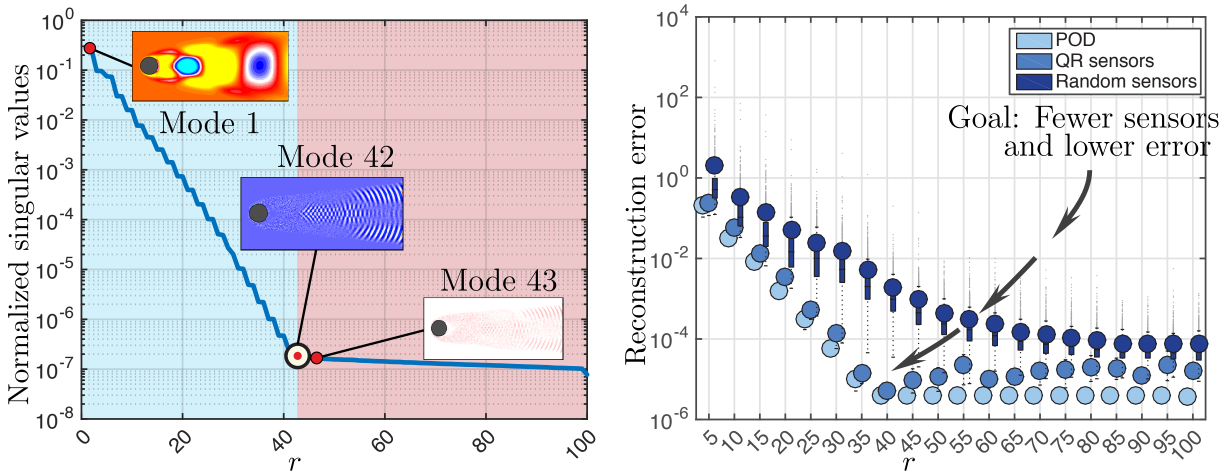


Figure 3.2: Fluid flow past a cylinder at Reynolds number 100, characterized by vortex shedding.

3.5.1 Flow past a cylinder

Fluid flow past a stationary cylinder is a canonical example in fluid dynamics that is high-dimensional yet reveals strongly periodic, low-rank phenomena. It is included here as an ideal system for reduction via POD and hence, minimal sensor placement. The data is generated by numerical simulation of the linearized Navier-Stokes equations using the immersed boundary projection method (IBPM) based on a fast multi-domain method [145, 50]. The computational domain consists of four nested grids so that the finest grid covers a domain of 9×4 cylinder diameters and the largest grid covers a domain of 72×32. Each grid has resolution 450×200, and the simulation consists of 151 timesteps with $\delta t = 0.02$. The Reynolds number is 100, and the flow is characterized by laminar periodic vortex shedding [120]. Vorticity field snapshots are shown in Fig. 3.2. In the cylinder flow and sea surface temperature examples, each snapshot $\mathbf{x}_i = \mathbf{x}(t_i)$ is a spatial measurement of the system at a given time t_i. Thus POD coefficients $a_k(t_i)$ are time dependent, and $\boldsymbol{\psi}_k(x)$ are spatial eigenmodes. The first 100 cylinder flow snapshots are used to train POD modes and QR sensors, and reconstruction error bars are plotted over 51 remaining validation snapshots in Figure 3.4.

The POD modes of this data reflect oscillatory dynamics characterized by periodic vortex-shedding. The data is low-rank, and the singular values decay rapidly, as shown in Fig. 3.3a. The singular values occur in pairs corresponding to harmonics of the dominant vortex shed-



(a) Singular values

(b) Reconstruction error with QR pivot sensors

Figure 3.3: Illustration of singular value decay and sparse signal reconstruction in the fluid flow past a cylinder example. (a) The exponential decay of singular values indicates low-rank dynamics with a modal truncation at $r = 42$ modes, the last modes containing any meaningful spatial structure. (b) Recovery with a minimal number of QR sensors with increasing POD basis rank is nearly as good as a snapshot's approximation in the same POD basis, and overfitting occurs beyond $r > 42$. Optimized sensing remains orders of magnitude more accurate than reconstruction with random sensors.

ding frequency. Most of the spectral energy in the dataset is captured by the first 42 POD modes. Thus the intrinsic rank of the dataset in POD feature space is $r = 42$, and the minimal number of QR pivots is $p = 42$. This yields a dramatic reduction from the initial state dimension of $n \approx 90000$ spatial gridpoints. Here, QR pivoting of $\mathbf{\Psi}_r^T$ (with $O(nr^2)$ operations) bypasses expensive $O(n^3)$ factorizations of large $n \times n$ matrices with alternate sampling or convex optimization methods.

Reconstruction from QR sensors (Fig. 3.3b) successfully captures modal content with only $p = r$ sensors when fitting to the first 42 POD modes. The first 42 POD modes characterize nearly 100% of the system's energy, the normalized sum of the singular values. Using $r > 42$ modes results in overfitting, and QR pivoting selects sensors based on uninformative modes. Thus, accuracy stops improving beyond $r = 42$ target modes, whereupon sensor data amplifies noise. However, these tailored sensors perform significantly better than random

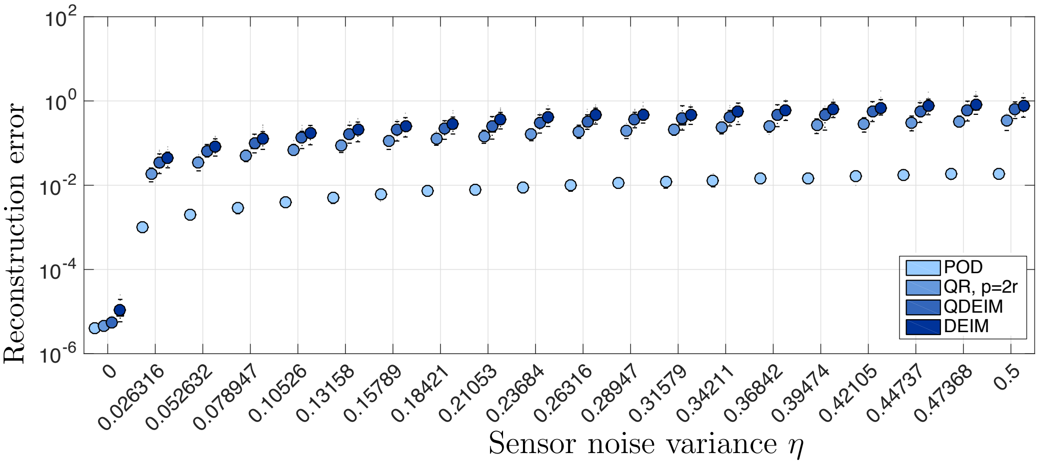


Figure 3.4: Comparison with DEIM. Oversampled QR permits additional sensors $p = 2r$ and a 4x reduction in reconstruction error compared to discrete empirical interpolation method (DEIM). In the $p = r = 40$ case DEIM and QR pivoting (Q-DEIM) perform comparably, with a slight improvement observed with QR. All sampling methods are within some constant factor of the optimal POD approximation with the full states. Hence POD-based sampling methods demonstrate robust, bounded growth of reconstruction error as sensor noise increases.

sensors due to the favorable conditioning properties of QR interpolation points.

3.5.2 Noise comparison study

Measurements of real-world data are often corrupted by sensor noise. The POD-based sensor selection criteria, as well as A,D and E-optimal experimental design criteria, are optimal for estimation with measurements corrupted by zero-mean Gaussian white noise. We empirically demonstrate this on the cylinder flow data with increasing additive white noise. Here we assume sensor noise only in the test measurements and not in the training data or features, see Eqn. (3.19). The POD modes and the different sensor sets are trained on the first 100 snapshots, and these different sensor sets are used to reconstruct the remaining 50 validation snapshots, which were not used for training features.

The reconstruction accuracy of the various sampling methods are compared for increasing sensor noise in Fig. 3.4, alongside the full-state POD approximation for illustration. Here we truncate the POD expansion to $r = 40$ eigenmodes and compare the $p = r$ reconstruction computed with the discrete empirical interpolation method (DEIM) [41] against the QR pivoting reconstruction (Q-DEIM, $p = r$). The DEIM greedy strategy places sensors at extrema of the residual computed from approximating the k-th mode with the previous $k-1$ sensors and eigenmodes. It can be seen that QR reconstruction is slightly more accurate than that of DEIM, which is the leading sampling method currently in use for reduced-order modeling [12].

QR pivoting is competitive in both speed and accuracy. The speed of QR pivoting is enabled by several implementation accelerations; for example, the column norms in line 4 of Algorithm 2 are only computed once and are then reused. Unlike QR pivoting, DEIM and related methods add successive sensors per iteration by similarly optimizing some metric over all locations. However, this metric (e.g., the approximation residual or the largest singular value) is recomputed at every iteration. The QR factorization is significantly faster than convex optimization methods used in optimal design of experiments, which typically require one matrix factorization per iteration.

In fact, convex optimization methods that relax the subset selection to weighted sensor placement provide no bounds for deviation from the global optimum, partly because rounding procedures are employed on the weights to decide the final selection. However, reconstruction error bounds for the globally optimal selection are known for DEIM [41], Q-DEIM [60] and related POD sampling methods [165, 171]. Furthermore, QR pivoting can achieve significant accuracy gains over DEIM by oversampling, i.e., when $p = 2r$ QR reconstruction error is 4x smaller than that of DEIM. It should be noted that while DEIM has not yet been extended to the $p > r$ case, oversampling methods exist for other POD-sampling methods [165, 171]. However, the iterative procedures involved in the latter are typically more expensive. Recent accelerated variants of greedy principled sampling [173] may permit oversampling for large n, when oversampled QR storage requirements would be excessive. In the cylinder flow case, we bypass this storage requirement by uniformly downsampling the fine grid by a factor of 5 in each spatial direction, thus reducing the number of candidate sensor locations to $n = 3600$

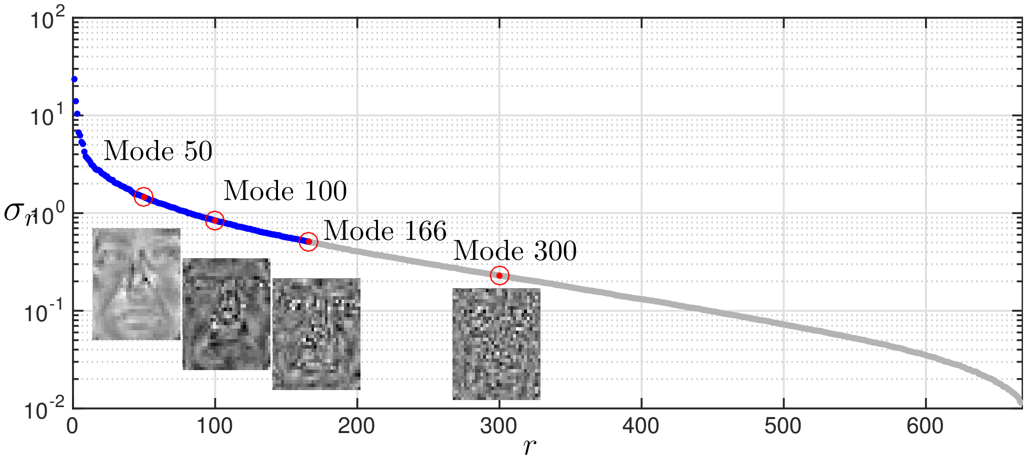


Figure 3.5: Normalized singular values and selected eigenfaces. Facial information progressively decreases across selected eigenfaces, and no facial features can be readily discerned beyond eigenface $r = 166$, the optimal modal truncation value according to [66].

instead of $n = 89351$.

3.5.3 Extended Yale B eigenfaces

Image processing and computer vision commonly involve high-resolution data with dimension determined by the number of pixels. Cameras and recording devices capture massive images with rapidly increasing pixel and temporal resolution. However, most pixel information in an image can be discarded for subject identification and automated decision-making tasks.

The extended Yale B face database [67, 102] is a canonical dataset used for facial recognition, and it is an ideal test bed for recovering low-rank structure from high-dimensional pixel space. The data consists of 64 aligned facial images each of 38 stationary individuals in different lighting conditions. We validate our sensor (pixel) selection strategy by recovering missing pixel data in a validation image using POD modes or *eigenfaces* trained on 32 randomly chosen images of each individual.

Normalized singular values are shown in Fig. 3.5, and the optimal singular value truncation threshold [66] occurs at $r = 166$, indicating the intrinsic rank of the training dataset.

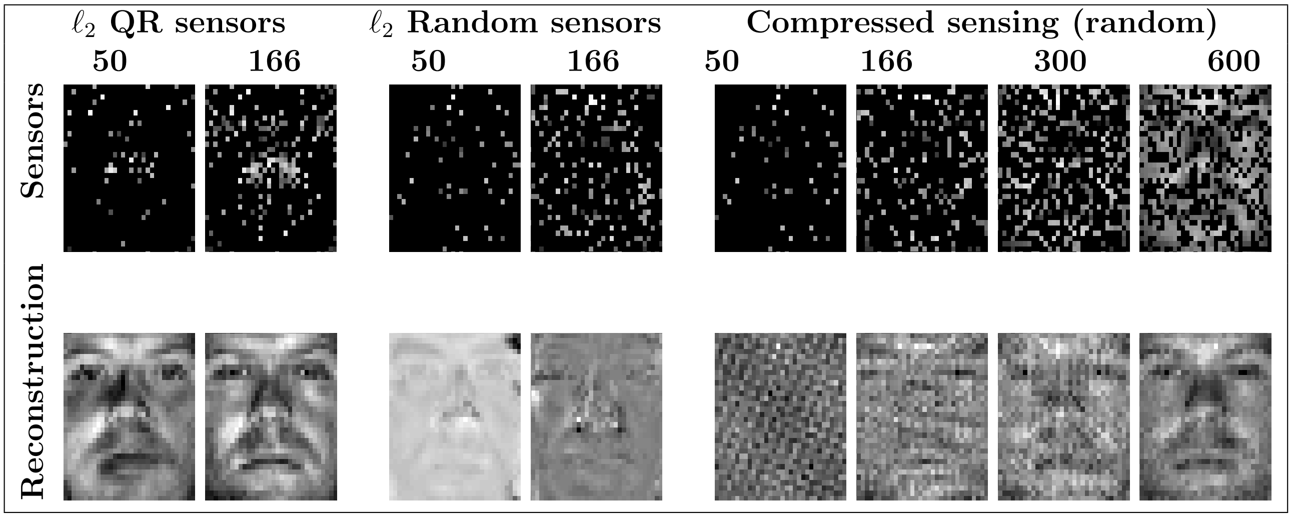


Figure 3.6: Comparison of reconstruction and sensor placement methods. QR sensors emphasize facial features such as the eyes, nose and mouth, and hence achieve adequate reconstruction with as few as 166 sensors (16% of all pixels). Least squares (ℓ_2) reconstruction with 50 QR sensors (5% of pixels) surpasses the performance of compressed sensing with 300 random pixels. In comparison, compressed sensing requires 600 sampled pixels for comparable recovery, while ℓ_2 reconstruction with random sensors is consistently poor.

Indeed, selected eigenfaces are also shown to reveal no meaningful facial structure beyond eigenface 166. QR pixel selection is performed on the first 50 and first 166 eigenfaces, and selected pixels shown in Fig. 3.6 cluster around important facial features – eyes, nose and mouth.

Image reconstructions in Fig. 3.6 are estimated from the same number of selected pixels as the number of modes used for reconstruction. For instance, the 50 eigenface reconstruction is uniquely constructed from 50 selected pixels out of 1024 total – 5% of available pixels. Even at lower pixel selection rates, least squares reconstruction from QR selected pixels is more successful at filling in missing data and recovering the subject's face.

For comparison, reconstruction of the same face from random pixels using compressed sensing is shown in Fig. 3.6. Compressed sensing in a universal Fourier basis demonstrates progressively improved global feature recovery. However, more than triple the pixels are

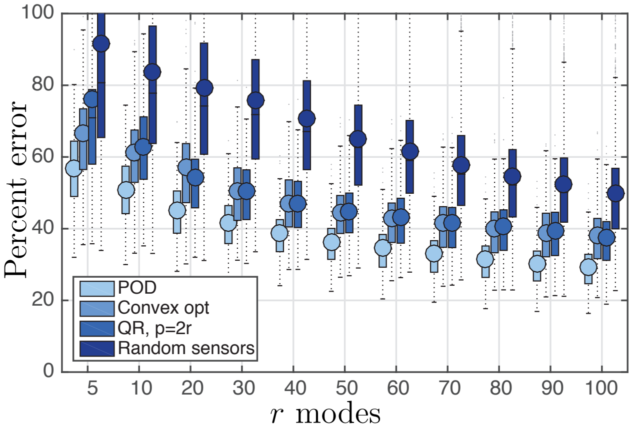


Figure 3.7: Recovery using our oversampled QR pivoting method for $p = 2r > r$ sensors. QR pivoting sensors perform comparably to sensors obtained via the convex optimization (Convex opt) method [84], and both are close to the optimal proper orthogonal decomposition approximation using the full state. In contrast to the convex method, QR sensors are obtained at reduced computational cost using the QR factorization of $\boldsymbol{\Psi}_r\boldsymbol{\Psi}_r^T$. Both methods outperform an equal number of randomly selected sensors.

required for the same quality of reconstruction as in QR selection. Moreover, the convex ℓ_1 optimization procedure is extremely expensive compared to the single ℓ_2 regression on subsampled eigenfaces. Therefore data-driven feature selection and structured measurement selection are of significant computational and predictive benefit, and occur at the small training expense of one SVD and pivoted QR operation.

The convergence of reconstruction with sensors using QR pivoting is shown in Fig. 3.7. More sensors than modes are used in reconstruction for this example. The expected error dropoff is observed with increasing number of modes and sensors, although the dropoff is slower than for the cylinder flow (Fig. 3.3b) due to slower decay of singular values.

3.5.4 Sea surface temperature (SST)

Next we consider the NOAA_OISST_V2 global ocean surface temperature dataset spanning the duration 1990–2016. The data is publicly available online [1]. Unlike eigenfaces, this

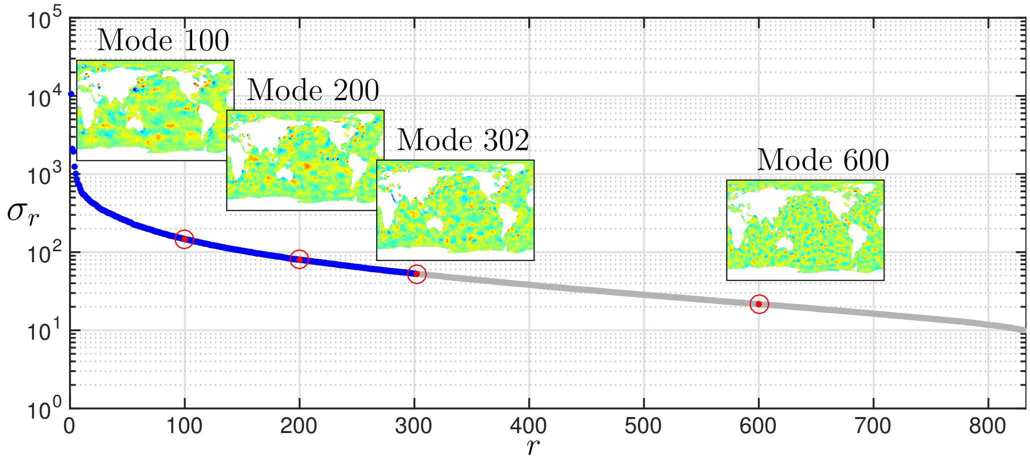


Figure 3.8: Singular values and selected proper orthogonal decomposition (POD) modes for sea surface temperature data. The optimal rank truncation threshold occurs at $r = 302$. POD mode 302 contains energetic localized convective phenomena (El Niño) but is largely uninformative on a global scale. Thus, overfitting may occur as more modes are included (see Fig. 3.10).

dataset is a time series, for which a snapshot is recorded every week. Sensor selection must then track energetic temporal signatures. Sensors and features are trained on the first 16 years (832 snapshots), and a test snapshot is selected from the excluded validation set. The singular values are shown in Fig. 3.8.

Like the eigenfaces, localized convective phenomena have energetic contributions to otherwise globally uninformative *eigenssts*. This is best seen in the POD snapshot projections, in which the 100 eigensst projection already sufficiently recovers dynamics, while increasing the number of eigenssts in the projection further refines convective phenomena. These lower-energy modes containing convective effects contribute to some degree of overfitting in ℓ_2 reconstruction (Fig. 5.7). The most interesting of these is the El Niño southern oscillation (ENSO) feature that is clearly identified from QR selected sensors. El Niño is defined as any temperature increase of a specified threshold lasting greater than six months in this highlighted region of the South Pacific. It has been implicated in global weather patterns, warming and climate change.

Remark: Modal separation of intermittent phenomena such as the El Niño is difficult from

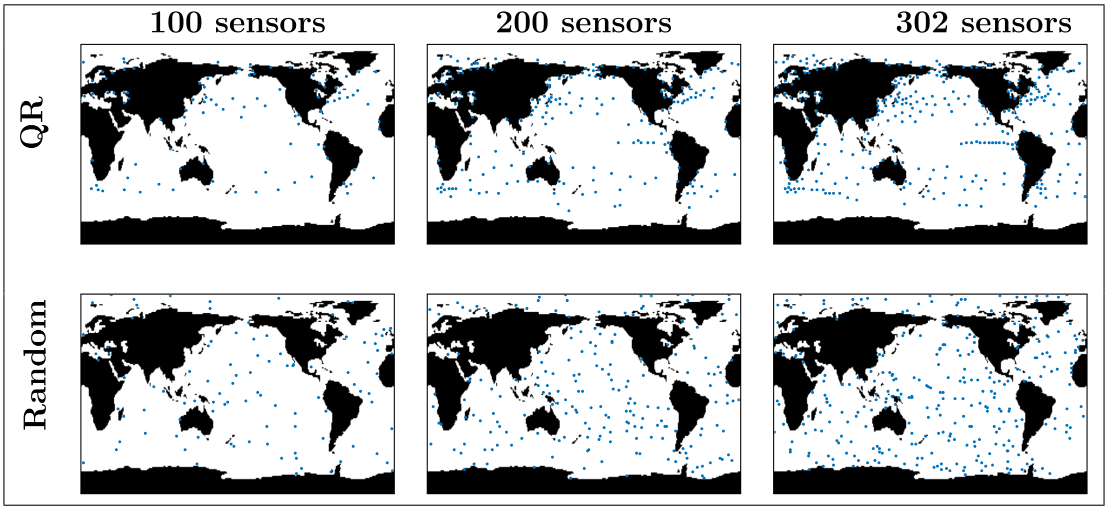


Figure 3.9: QR selected sensors used for reconstruction. QR sensors are informative about ocean dynamics, for example capturing convective phenomena such as the El Niño Southern Oscillation off coastal Peru.

a time-invariant POD analysis. Separation of isolated, low-energy temporal events cannot be done from a variance-characterizing decomposition such as the POD – reordering the snapshots will yield the same dominant modes. On the other hand, tensor decompositions and temporal-frequency analyses such as multiresolution dynamic mode decomposition have succeeded at identifying El Niño where POD has failed. Sensor selection using non-normal modes arising from such decompositions remains an open problem and the focus of ongoing work.

3.6 Conclusions

In this section, we have demonstrated the practical implementation of several sparse sensing algorithms on a number of relevant real-world examples. As discussed throughout, there is no all-purpose strategy for the sparse sensing of a high-dimensional system. Instead, the choice depends on key factors such as the amount of training data available, the scope and focus of the desired estimation task, cost constraints on the sensors themselves, and the required latency of computations on sensor data. Thus, we partition the sparse sensing algorithms

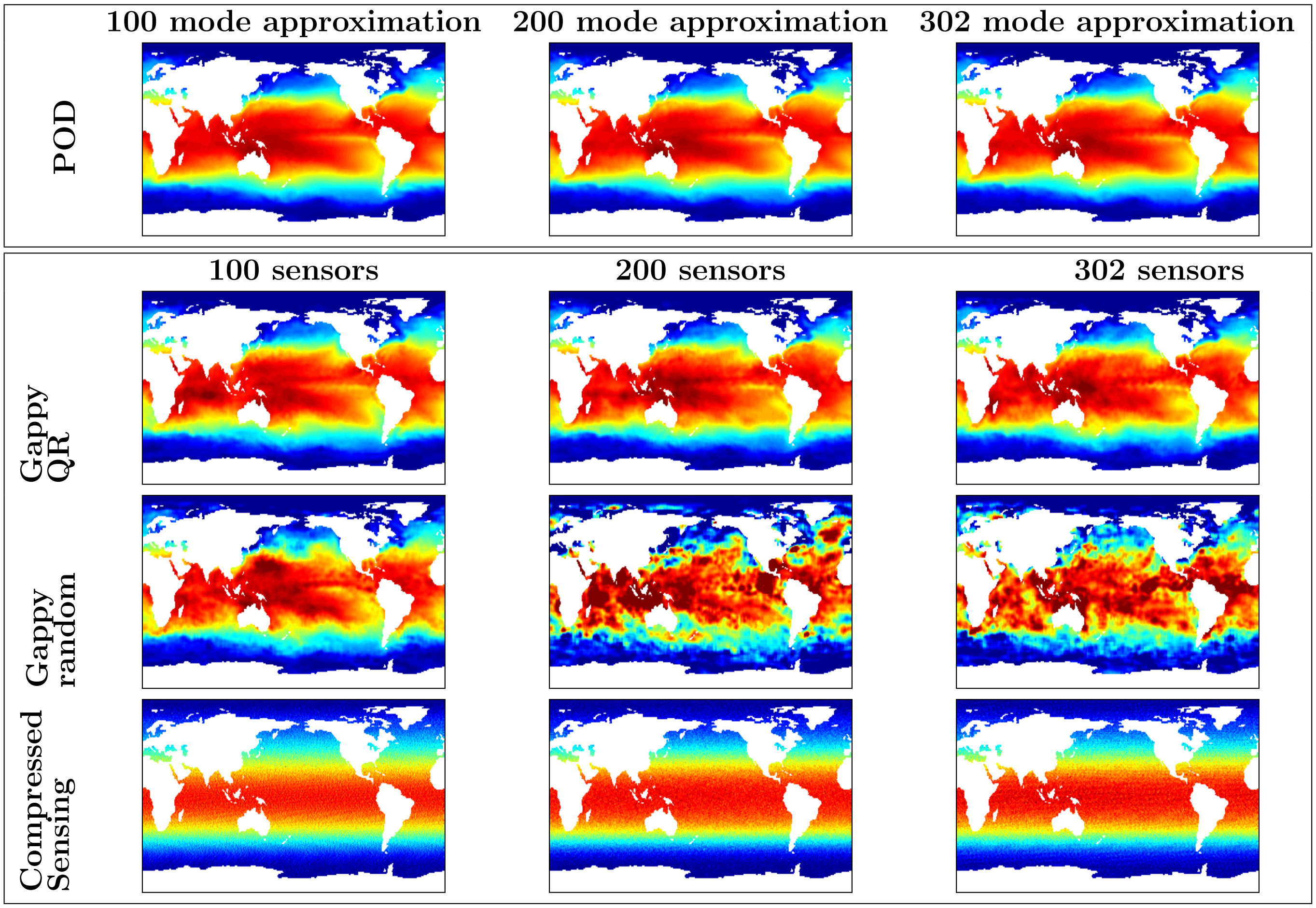


Figure 3.10: Comparison of methods for reconstruction of a single snapshot from sensors. QR selected sensors filter uninformative features and achieve better reconstruction. In comparison, random sensors achieve poor reconstruction with least squares (ℓ_2) and compressed sensing. As the number of proper orthogonal decomposition (POD) modes is increased, modal approximation with the full state (top row) only gets marginally better, which indicates additional features contribute low-energy uninformative features. Hence ℓ_2 reconstruction expresses low-energy POD modes and suffers from overfitting, with both random and to a lesser extent, QR sensors.

into two fundamental categories: 1) optimized sensing in a data-driven tailored basis, and 2) random sensing in a universal basis.

A critical comparison of the two approaches highlights a number of relative strengths and weaknesses. The first strategy results in a highly optimized set of sensors that are suitable for tightly scoped reconstruction problems where sufficient training data is available. The second strategy requires more sensors for accurate reconstruction but also makes fewer assumptions about the underlying signal, making it more general. We emphasize that optimized sensing in a tailored basis typically provides more accurate signal reconstruction than random measurements, facilitating a reduction in the number of sensors by about a factor of two. Further, sensor selection and signal reconstruction in the tailored basis is computationally efficient and simple to implement, while compressed sensing generally requires a costly iterative algorithm. In problems where sensors are expensive, or when low-latency decisions are required, the reduction in the number of sensors and the associated speed-up of optimized sensing can be significant. Thus, when the reconstruction task is well-scoped and a sufficient volume of training data is available, we advocate principled sensor selection rather than compressed sensing. In addition, pivoted QR sensors may be used in conjunction with other tailored bases (polynomials, radial basis functions) when signal structure is known. Since these are not data-driven basis functions, QR optimized samples can generalize to different dynamical regimes or flow geometries.

3.6.1 Potential applied impact

Many fields in science and engineering rely on sensing and imaging. Moreover, any application involving feedback control for stabilization, performance enhancement, or disturbance rejection relies critically on the choice of sensors. We may roughly categorize these sensor-critical problems into two broad categories: 1) problems where sensors are expensive and few (ocean sampling, disease monitoring, espionage, etc.), and 2) problems where sensors are cheap and abundant (cameras, high-performance computation, etc.).

In the first category, where sensors come at a high cost, the need for optimized sparse sensors is clear. However, it is not always obvious how to collect the training data required to optimize these sensors. In some applications, high-fidelity simulations may provide insight into coherent structures, whereas in other cases a large-scale survey may be required. It has recently been shown that it may be possible to optimize sensors based on heavily subsampled data, as long as coherent structures are non-localized [23].

In the second category, where sensors are readily available, it may still be advantageous to identify key sensors for fast control decisions. For example, in mobile applications, such as vision-based control of a quad-rotor or underwater monitoring of an energy harvesting site using an autonomous underwater vehicle, computational and battery resources may be limited. Restricting high-dimensional measurements to a small subset of key pixels speeds up computation and reduces power consumption. Similar performance enhancements are already exploited in high-performance computing, where expensive function evaluations are avoided by sampling at key interpolation points [42, 41]. Finally, it may also be the case that if measurements are corrupted by noise, reconstruction may improve if uninformative sensors are selectively ignored.

Chapter 4

SENSORS FOR PREDICTIVE MANUFACTURING

Much of the data collected from real-world processes contains outliers and noise that may skew features resulting from PCA. In these cases features *robust* to outliers and noise are desirable in practice. Indeed, robust statistical methods are critically important for evaluating real-world data, as advocated by John W. Tukey in the earliest days of data science [81, 59]. This chapter investigates an application to predictive aircraft manufacturing, in which we combine robust dimensionality reduction and sparse measurement optimization to dramatically reduce the number of measurements required to shim gaps between assembled parts in a modern aircraft. The first section motivates the data-driven *predictive shimming* of an aircraft, followed by a description of feature selection using robust principal component analysis. These robust principal component analysis features serve as a proxy for standard POD modes in the data-driven learning and QR pivoting sensor placement step. The final section demonstrates successful prediction using our method for gaps occurring in the wing to body join across 54 representative Boeing aircraft.

4.1 Motivation

Aircraft are built to exceedingly high tolerances, with components sourced from around the globe. Even when parts are manufactured to specification, there may be significant gaps between structural components upon assembly. One of the most time-consuming and expensive efforts in part-to-part assembly is the shimming required to bring an aircraft into the engineering nominal shape. Historically, parts have been dry-fit, gaps measured manually, and custom shims manufactured and inserted, often involving disassembly and reassembly. Recent advancements in 3-D scanning have enabled their use for surface measurement prior

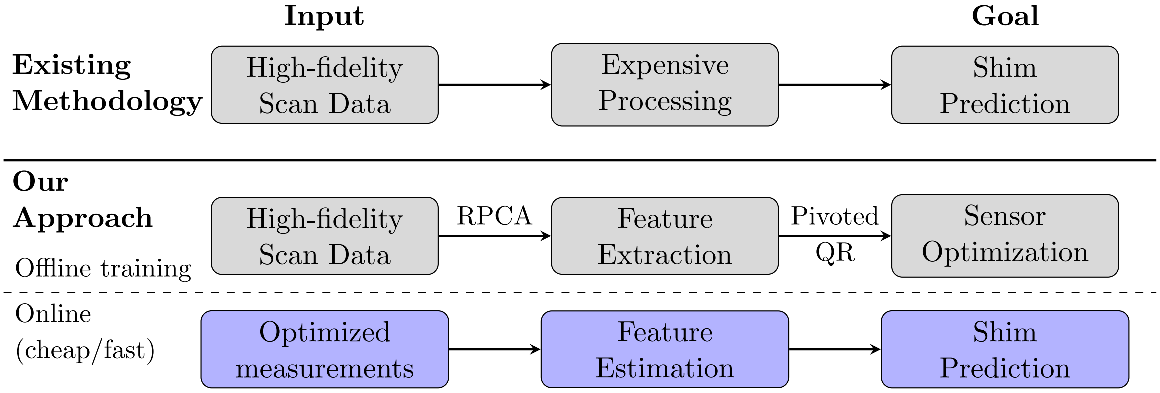


Figure 4.1: **Schematic overview of predictive shimming.** Sparse sensor optimization is used to accelerate predictive shimming of aircraft. We compute robust shim gap features and measurement locations in a offline training stage. This dimensionality reduction in the offline training stage facilitates an extremely low-dimensional estimation problem to recover high-fidelity shims during online execution. Note that high-fidelity measurements are never needed online.

to assembly, known as *predictive shimming* [83, 117, 115, 45, 116, 114]. There are several patents and papers describing methods of high-tolerance measurement and manufacturing required for predictive shimming [107, 108, 45, 19, 162, 159, 20, 5]. While some cost-effective devices may not provide the fidelity required, higher-fidelity metrology and scanning devices result in overwhelming amounts of data, shifting the burden from time-consuming manual shimming to time-consuming computational processing for predictive shimming. This amounts to significant delays in production, with much of the time being spent in the critical path of assembly. Reducing the burden of data collection and processing, and ultimately reducing delays for optimized aircraft assembly, could have significant financial implications, on the order of billions of dollars a year.

Our framework for data-driven measurement placement suggests an alternative approach to predictive shimming. Instead of measuring each component of a new aircraft in isolation, we leverage historical production data to learn low-rank patterns from the spatial shim gap distributions. The robust principal component analysis (RPCA) [39] is used to compute low-rank gap principal components that are robust to outlier measurements. RPCA is based on

Algorithm 3 Inexact augmented Lagrange multiplier method for Robust PCA
Usage: INEXACT_ALM_RPCA($\mathbf{X}$, λ)

```
procedure INEXACT_ALM_RPCA( $\mathbf{X}$, $\lambda$ )
    Initialize $\mathbf{X}, \mathbf{L}, \mathbf{S}, \mu$
    while not converged do
        minimize $\mathcal{L}(\mathbf{L}, \mathbf{S}, \mathbf{Y})$ by updating $\mathbf{L}$ and $\mathbf{S}$ only once
        $\mathbf{Y} = \mathbf{Y} + \mu(\mathbf{X} - \mathbf{L} - \mathbf{S})$
        update $\mu$
return $\mathbf{L}, \mathbf{S}$
```

the computationally efficient singular value decomposition (SVD) [72], and yields the most correlated spatial structures in the aircraft measurements, identifying areas of high variance across different aircraft. Next, QR pivoting is applied to these RPCA modes to obtain a minimal subset of key measurement locations (sensors) that are most predictive of the spatial shim gap distribution. We apply our algorithms to a historical production dataset consisting of 54 Boeing aircraft. The optimal measurements determined by our method successfully predict 99% of the shim gaps within the desired measurement tolerance using approximately 3% of the available laser scan data. All results are rigorously cross-validated.

4.2 Robust Principal Component Analysis

Least-squares regression is highly susceptible to outliers and corrupted data. Principal component analysis (PCA) suffers from the same weakness, making it *fragile* with respect to outliers. To address this sensitivity, [39] introduced a robust principal components analysis (RPCA) that decomposes a data matrix $\mathbf{X}$ into a low-rank matrix $\mathbf{L}$ containing dominant coherent structures, and a sparse matrix $\mathbf{S}$ containing outliers and corrupt data:

$$\mathbf{X} = \mathbf{L} + \mathbf{S}. \tag{4.1}$$

The principal components of $\mathbf{L}$ are *robust* to the outliers and corrupt data in $\mathbf{S}$. The RPCA decomposition has tremendous applicability for many modern problems of interest, including video surveillance [17] (where the background objects appear in **L** and foreground objects appear in **S**), natural language processing [92], matrix completion and face recognition [170].

Mathematically, the goal of RPCA is to find matrices **L** and **S** that satisfy

$$\min_{\mathbf{L},\mathbf{S}} \operatorname{rank}(\mathbf{L}) + \|\mathbf{S}\|_0 \text{ such that } \mathbf{L}+\mathbf{S}=\mathbf{X}. \tag{4.2}$$

However, neither the rank(**L**) nor the $\|\mathbf{S}\|_0$ terms are convex, and this is not a scalable optimization problem. Similar to compressed sensing, it is possible to solve for the optimal **L** and **S** with *high probability* using a convex relaxation of (4.2):

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \text{ such that } \mathbf{L}+\mathbf{S}=\mathbf{X}. \tag{4.3}$$

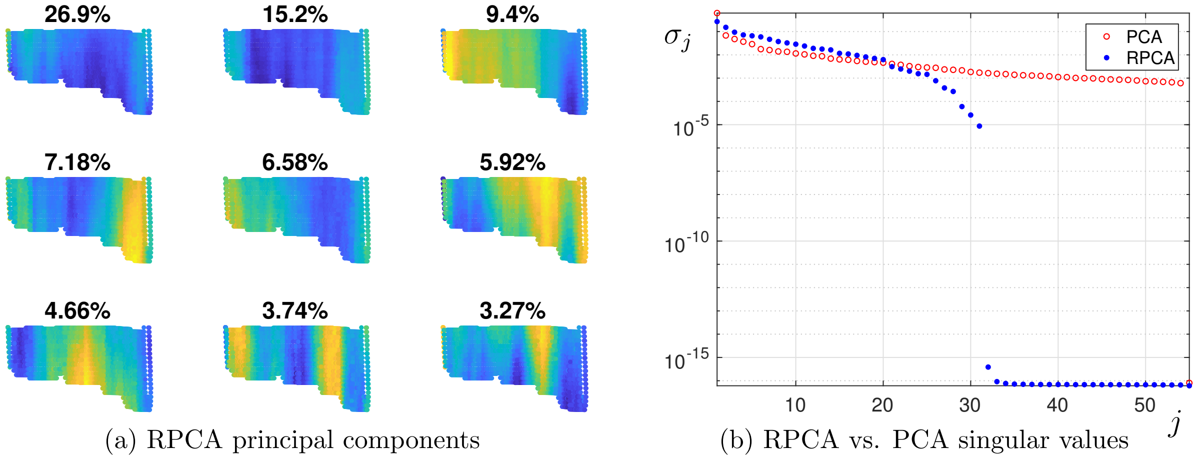
Here, $\|\cdot\|_*$ denotes the nuclear norm, given by the sum of singular values, which is a proxy for rank. The solution to (4.3) converges to the solution of (4.2) with high probability if $\lambda = 1/\sqrt{\max(n,m)}$, where n and m are the dimensions of **X**, given that **L** is low-rank and **S** is sparse.

The problem in (4.2) is known as *principal component pursuit* (PCP), and may be solved using the augmented Lagrange multiplier (ALM) algorithm. The augmented Lagrangian may be constructed as:

$$\mathcal{L}(\mathbf{L},\mathbf{S},\mathbf{Y}) = \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 + \langle \mathbf{Y}, \mathbf{X}-\mathbf{L}-\mathbf{S}\rangle + \frac{\mu}{2}\|\mathbf{X}-\mathbf{L}-\mathbf{S}\|_F^2. \tag{4.4}$$

A general solution would solve for the $\mathbf{L}_k$ and $\mathbf{S}_k$ that minimize $\mathcal{L}$, update the Lagrange multipliers $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu(\mathbf{X}-\mathbf{L}_k-\mathbf{S}_k)$, and iterate until the solution converges. This is outlined in Algorithm 3. For this specific system, the alternating directions method (ADM) [**?**, 172] provides a simple procedure to find **L** and **S**. The parameter μ is discussed more in [172] and [39].

In the following, RPCA will be used to develop low-dimensional representations for high-dimensional aircraft metrology data (e.g., laser scans or point cloud measurements). In



(a) RPCA principal components (b) RPCA vs. PCA singular values

Figure 4.2: (a) **Robust principal components of shim gaps**. This view is a projection orthonormal to one of the mating surfaces, where color corresponds to principal components of measured gap thickness from small (blue) to large (red). Here robust principal components 1-9 are displayed alongside the percentage of the total variance captured in each component, which also reveal segmented regions of spatial features that correspond to individual shim parts (Figure 4.5). (b) **Normalized singular values.** Singular values, which encode the percentage of the total variance contained in each principal component, indicate that nearly all the variance in the shim gap data is contained in the first 31 RPCA principal components. By contrast, the slow decay of PCA singular values indicates that standard PCA features are corrupted by noise which would contribute to overfitting validation data.

particular, the left singular vectors $\mathbf{\Psi}$ of the low-rank matrix $\mathbf{L}$ yield robust principal components, and are computed via the SVD:

$$\mathbf{L} = \mathbf{\Psi}\mathbf{\Sigma}\mathbf{V}^*. \tag{4.5}$$

These low-rank coherent patterns will then facilitate sparse sensing strategies.

4.3 Measurement optimization and gap prediction

Assuming that shim gaps are well characterized by r low-rank features $\mathbf{\Psi}_r$, we seek the optimal $p \geq r$ gap measurement locations that best characterize the remaining $n - p$ shim gaps in a procedure identical to that presented in Chapter 3. As before, the optimal sensor measurements are computed from the low-rank RPCA modes using Algorithm 2

$$\Psi_r^T \mathbf{C}^T = \mathbf{QR}, \tag{4.6}$$

or, when $p > r$,

$$\boldsymbol{\Psi}_r \boldsymbol{\Psi}_r^T \cdot \mathbf{C}^T = \mathbf{QR}. \tag{4.7}$$

Subsequently the $n - p$ unknown shim gaps for a new aircraft are predicted from the p measured gap values at the optimal locations stored in, $\mathbf{y} = \mathbf{Cx}$, with gappy RPCA reconstruction

$$\hat{\mathbf{x}} = \boldsymbol{\Psi}_r (\mathbf{C}\boldsymbol{\Psi}_r)^\dagger \mathbf{y} \tag{4.8}$$

These optimal locations and features are obtained in the training stage using the QR column pivoting method and RPCA methods respectively, effectively selecting measurement locations from regions of strong spatial correlation. Finally, we assess the success of the prediction with the percentage of points that differ from the true gap values by less than 0.005”, which is the desired measurement tolerance for a laser scanner, or equivalently, the number of indices i out of n total that satisfy $|x_i - \hat{x}_i| < 0.005$.

Finally, it is useful to decompose a large aircraft structure into several smaller shim components, which are each analyzed independently. As we see, this improves the prediction performance and reduces the number of measurements by ensuring that the data remains tightly correlated across aircraft.

4.4 Results and discussion

We demonstrate the performance of the proposed architecture on production data from a challenging part-to-part assembly on a Boeing aircraft. This data set consists of 10076 laser gap measurements of the part assembly for 54 different production instances of the same aircraft type. Measurement locations are aligned between the instances, making the data amenable to SVD. We build a reduced-order model of the shim distribution using RPCA and

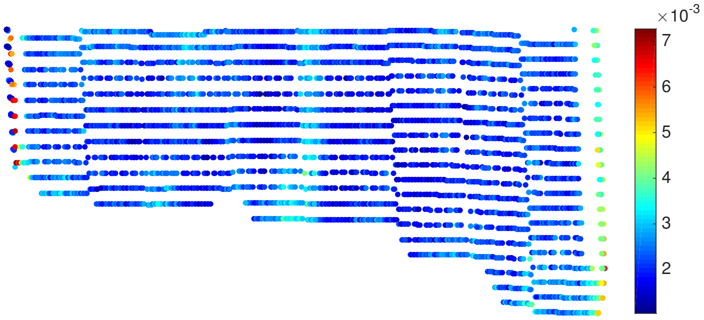


Figure 4.3: **Median gap prediction error**. The median prediction error at almost all spatial locations falls below the manufacturing measurement tolerance of 0.005 inches. The median at each point is taken across 53 instances of cross-validated prediction. The subsequent figure plots the probability distribution of prediction error.

then design optimized measurement locations based on these data-driven features. We train the model on 53 aircraft and then validate on the remaining aircraft; this process is repeated for all 54 possible training/test partitions to cross-validate the results. Thus, a data matrix $\mathbf{X} \in \mathbb{R}^{10076 \times 53}$ of training data is constructed, in which each column contains all of the shim gaps for one aircraft, and each row contains the measured gap values at one specific location for all aircraft in the training set.

The low-rank decomposition of $\mathbf{X}$ via RPCA is shown in Fig. 4.2. The singular value distribution indicates dominant low-rank structure, and the principal components illustrate the coherent correlated patterns observed in the shim gap data. RPCA has a tunable parameter λ that controls the strength of outlier rejection or the number and magnitude of nonzero entries in $\mathbf{S}$. In practice, λ is set to an optimal value determined by the dimensions of the matrix, which in our case is $\lambda = 1/\sqrt{\max(10076, 53)}$.

4.4.1 Prediction results for entire wing to body join

First, we investigate sensor optimization and shim prediction considering the entire data set as single large structure. In the next section, we decompose the data into several individual shim segments.

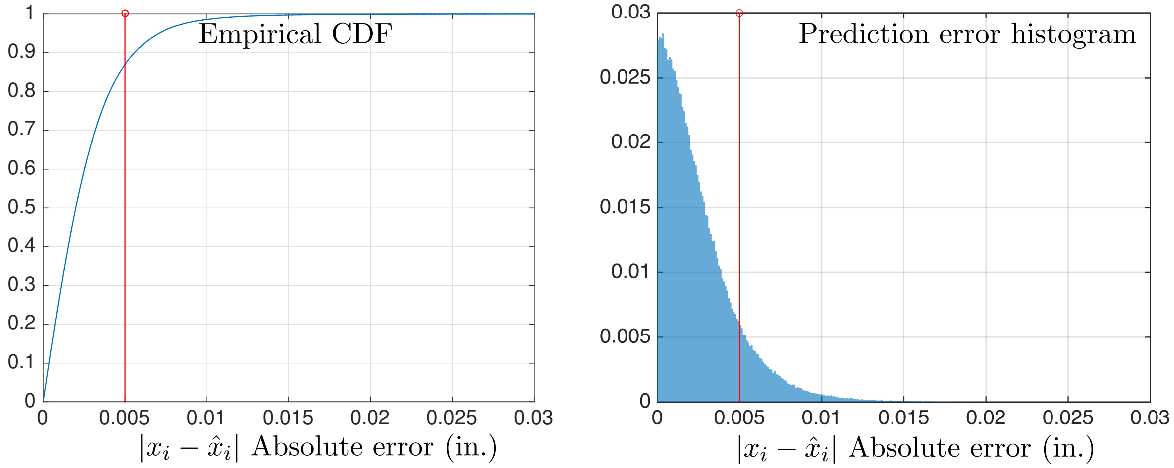
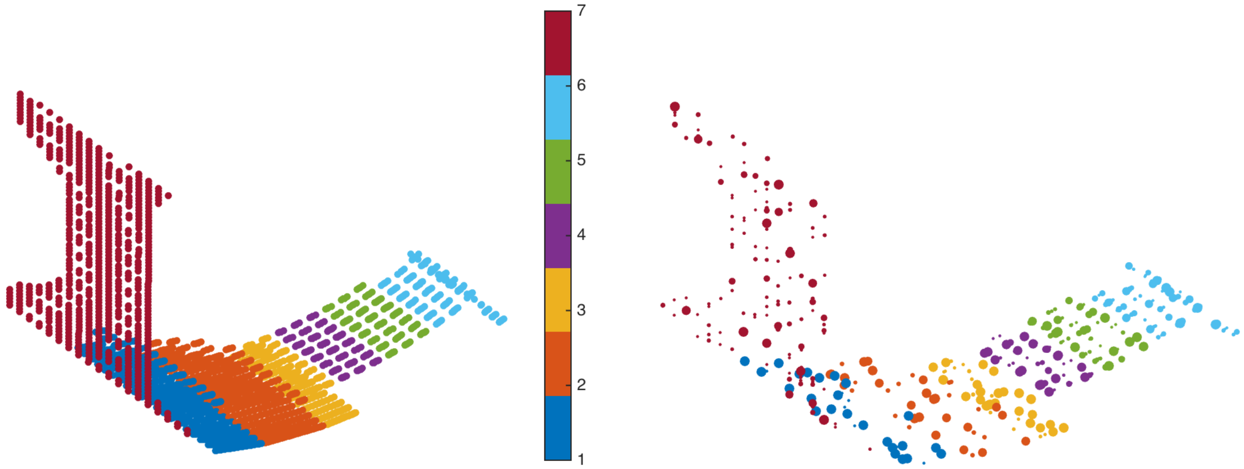


Figure 4.4: **Pointwise absolute error distribution**. The normalized histogram of absolute error across 534,028 total validation points (10076 gaps × 53 aircraft) is shown on the right. The cumulative distribution function (CDF) of this distribution (left) shows that 87% of the predicted gap absolute errors are less than desired tolerance of 0.005 inches, indicated by the red threshold.

Our algorithm identifies $r = 31$ RPCA modes and $p = 100$ optimized measurements using oversampled QR pivoting. Potential overfitting may occur when fitting more than 100 measurements to these 31 features, since some measurements may be contaminated by noise not reflected in the feature space. Thus, increasing the number of measurements further does not yield substantial improvements in prediction accuracy.

The 100 optimized measurements consist of less than 1% of the original 10076 gap measurements. From these few measurements, the remaining shim gaps are predicted with high accuracy. Figure 4.3 shows the median pointwise absolute error between the predicted and true shim values. The prediction is successful with high probability across the 53 cross-validation cases. The absolute error distribution for all points across all tests, shown in Figure 4.4, reveals that 87% of all points are predicted within ±0.005". The few gap locations with error larger than the required ±0.005" will likely not contribute significantly to a manufactured shim, as these are averaged out in splining and fabrication. The few points that are outside of this prediction tolerance will be outweighed by the numerous successfully



(a) High-fidelity scan locations (b) Optimal measurement sensor ensembles

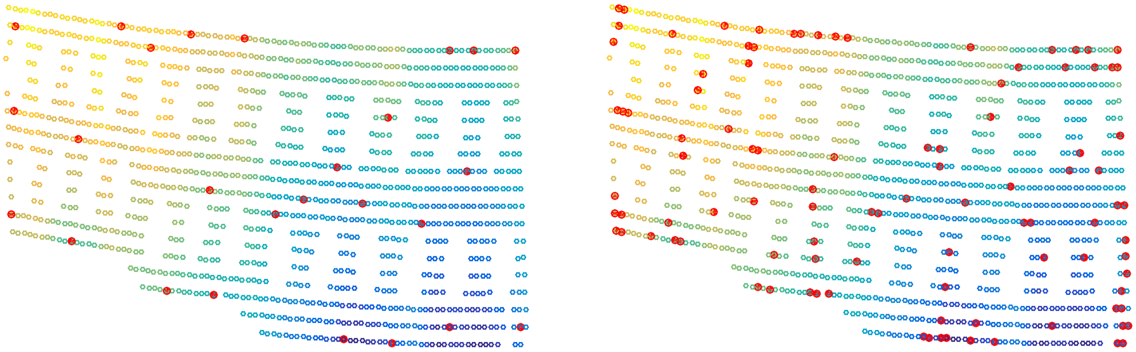
Figure 4.5: **Gap measurements**. This figure plots high-fidelity gap measurement locations segmented into 7 canonical shim regions (left). Ensembles of measurement locations (sensors) selected by our method are pictured on the right. Sensors are sized by the number of times they are selected by our method across 53 different gap measurement training sets. They are more concentrated around edges and corners, which is where we expect more systemic gap variation.

predicted gaps in the fabrication of the shim, as around 1000 laser scan points are used to define each shim.

4.4.2 Prediction results for segmented shims

The number of incorrectly predicted gaps can be reduced by training our method on each individual shim part separately, as there are seven individual shim segments in the assembly considered. Considering each shim segment separately helps to ensure that data is tightly correlated across aircraft, improving prediction. It is possible to reduce the number of total measurements and improve the overall performance using this segmentation approach.

Figure 4.5 displays the seven separately manufactured shim segments, as well as the sensor ensembles for each shim. Prediction results are shown in Table 4.1. Prediction accuracy is vastly improved, and 96-99% of the shim gap locations are predicted to within the desired



(a) 26 sensors (b) 100 sensors

Figure 4.6: **Optimal sensors for single segment.** Optimal measurements from our method (red) are plotted over gap data (yellow corresponds to larger gap sizes) for shim segment #2 (Fig. 4.5). (left) 26 optimal measurement locations required to predict all unmeasured gaps with 98.05% accuracy (Table 4.1). Since this shim is well-described by as few as $r = 26$ principal components, the minimum required number of optimal sensors is 26. Our method can also increase the number of optimal sensors (right), which increases sensor density near the previously selected locations. Although additional sensors provide redundant information, it is shown in Fig. 4.7 that they yield only marginal improvements in accuracy.

±0.005” tolerance. Furthermore, we note that that the rates of optimal measurements r vary from anywhere between 2% to 6% of all points within the shim, which indicates that some shims are higher-dimensional and require more features (hence, sensors) to be fully characterized. This is also reflected in the sensor ensembles in Fig. 4.5.

Figure 4.6 shows the optimized sensor locations for the second shim segment. As the number of sensors is increased from $p = 26$ to $p = 100$, the sensor density increases near

Shim No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Percent Accurate	97.90	98.05	99.82	99.94	99.99	99.03	99.97
Optimal sensors (avg)	26	26	25	26	25	26	25
Total points	1003	1116	453	692	709	768	664

Table 4.1: Segmented prediction results show vastly improved prediction accuracies, with 97-99% of gaps predicted to within the desired 0.005 inch measurement tolerance.

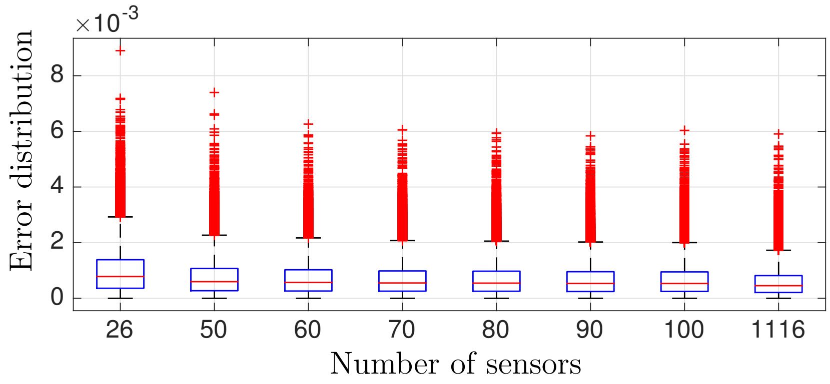


Figure 4.7: **Pointwise error vs. number of optimal sensors.** As the number of sensors are increased, pointwise error distributions do not show much improvement beyond $p = 60$ sensors for shim 2, which has 1116 points in total. For instance, at $p = 80$ only 3 points are mispredicted with absolute error greater than 0.005 inches out of all 1116 × 53 test points. Finally, the best prediction using all 1116 points yields only marginal improvements.

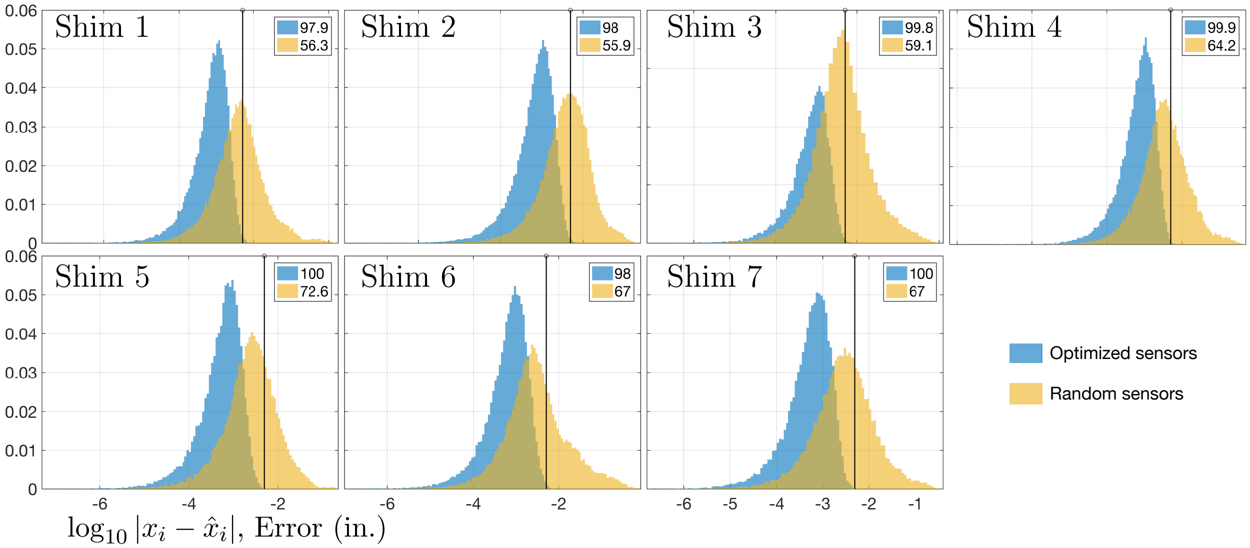


Figure 4.8: **Absolute error distribution for optimal (blue) vs. random (orange) sensors.** The above are histograms of pointwise absolute error across all validation tests using reconstruction from optimal and an equal number of *random* sensors (which are randomly selected from available measurement locations shown in Figure 4.5). The red line represents the desired measurement tolerance of 0.005 inches, and the legend indicates the percentage of points that fall within this error tolerance. The histograms indicate that optimal sensors predict nearly twice as well as random sensors.

the originally identified points. Figure 4.7 shows the error distribution as the number of sensors is increased, indicating that there is a performance cap at a certain small number of sensors. Finally, Fig. 4.8 shows the error distributions of predicted shim gaps for all seven segments, comparing the performance of optimized sensors versus randomly chosen measurement locations. In all cases, optimized sensors result in significantly more accurate shim prediction over random sensors, with the vast majority of predicted shims within the desired tolerance.

4.5 Conclusions

This work demonstrates the ability of data-driven sensor optimization to dramatically reduce the number of measurements required to accurately predict shim gaps in aircraft assembly. We combine sparse sensor optimization with robust feature extraction and apply the technique to historical Boeing production data from a representative aircraft. With around 3% of the original laser scan measurements, our algorithm is able to predict the vast majority of gap values to within the desired measurement tolerance, resulting in accurate shim prediction. These optimized measurements exhibit excellent cross-validated performance and may inform targeted, localized laser scans in the future. Reducing the burden of data acquisition and downstream computations has tremendous potential to improve the efficiency of predictive shimming applications, which are ubiquitous and often take place in the critical path of aircraft assembly. Thus, streamlining this process may result in billions of dollars of savings.

The proposed algorithm has been demonstrated on laser point scan data, although the method is broadly applicable to data from various metrology equipment. In the current study, reducing the number of laser scan points directly reduces measurement time, which is considerable. However, in many predictive shimming applications, vast and increasing volumes of point cloud data are being collected. As opposed to laser scans, point clouds are relatively inexpensive to acquire, and it is less clear that sparse sensing will significantly reduce acquisition time. Nonetheless, point cloud scans often yield millions or billions of data points, which result in computationally expensive downstream predictions. The sparse sensor algorithm may pinpoint key locations in this vast data, dramatically reducing shim computations, likewise reducing the time required for predictive shimming. In addition, because high-resolution scan data is available, it may be possible to use thousands of nearby points in the vicinity of an identified sparse sensor location to obtain a robust gap estimate in this region. It may also be possible to obtain *super-resolution*, leveraging the availability of large quantities of nearby data.

There are a number of interesting future directions for this work. In the present study, we use RPCA to remove outliers in the data, providing a robust computation of low-rank patterns. However, the sparse outliers may contain information about anomalies and defects, and machine learning may be used to automatically classify these outliers. It has also been observed in this data that aircraft tend to cluster in the PCA feature space, indicating possible differences in manufacturing processes across line number. Mining this data and using these PCA coefficients as diagnostics for process control may be possible.

Although the present work has been applied to historical production data, this technology can be implemented for online learning in aircraft production. Importantly, every step in the algorithm is relatively inexpensive, involving laptop computations on the order of minutes. The initial RPCA and pivoted QR are one-time computations, which can be accelerated by GPU implementations when datasets contain on the order of millions of points. The dimensionality reduction provided by RPCA permit each aircraft's gap data to be compactly represented by its PCA coefficients, plus a small number of PCA features shared by all aircraft. Then, the full data can be cheaply recovered from a simple projection onto PCA features. Thus the PCA framework (even without optimal measurements) permits compact storage of large data that often result from scans. After the offline sensor optimization, online predictive shimming merely requires 1) collecting a small subset of optimized measurements and 2) performing least-squares regression. The features and predictions only improve with increasing data, and so as more aircraft are manufactured, the dominant correlation structures, and resulting sparse sensor locations, become more clear.

Chapter 5

OPTIMIZED SAMPLING FOR MULTISCALE DYNAMICS

The POD reduced characterization of data is variance or energy based, which means that reordering time snapshots within the data matrix produces the same results. In addition to being dynamically irrelevant, POD modes miss rare temporal events that are typically not energetic in the system. Many natural systems exhibit *multiscale* transient, intermittent, or rare bursting dynamics that are not amenable to this approach. Indeed, the dynamic interactions that persist across multiple timescales are typical of many complex systems in fluid dynamics, material science, atmospheric and ocean interactions, networked systems and neurobiology. Dimensionality reduction techniques show immense promise in obtaining interpretable models directly from data for downstream prediction and control. However, complex interactions of spatiotemporal scales and limited measurement capabilities confound efforts to separate the various micro- and macro-scale physics. To alleviate this critical limitation, we develop a multiresolution analysis (MRA) [54, 56] algorithm for the separation of multiscale, low-rank phenomena by its intrinsic timescale in order to determine optimal measurement locations. We build on the multiresolution dynamic mode decomposition (mrDMD) [100] which capitalizes on a wavelet-like decomposition for MRA and embeds the dynamics at each timescale using a low-rank feature extraction technique such as POD.

The workhorse algorithm behind mrDMD is the dynamic mode decomposition (DMD) [137, 133, 157], a matrix decomposition that identifies low-rank spatial structures and their corresponding temporal Fourier dynamics [99]. In contrast to POD, DMD can be considered a combination of the Discrete Fourier Transform and POD for preserving both temporal correlation and frequency information. It further provides interpretable, low-rank features in data that can directly be used for prediction or for building dynamic models.

The low-rank patterns of DMD can be further exploited for sparse (greedy) sampling of the dynamics. Thus our secondary goal for multiscale characterization is to determine optimal sensor placement that respects the multiscale dynamics induced by the measured system. We propose applying QR pivoting directly to mrDMD modes for spatially sampling the multiscale phenomena of interest. This method accurately identifies dynamics occurring at different timescales in multiscale data. Secondly, multiscale sensors derived from QR matrix pivots of mrDMD modes are shown to spatially cover localized coherent structures exhibiting specific frequencies. We then leverage resulting mrDMD modes and multiscale sensors to estimate temporal behavior, for which multiscale reductions are demonstrably more accurate than POD-based reduction.

5.1 Background

This section develops the background theory of dynamic mode decomposition and its multiresolution variant mrDMD. Our goal is to construct a spatiotemporal decomposition directly from time snapshots of high-dimensional system states $\mathbf{x} \in \mathbb{R}^n$, such that their evolution can be expressed as a linear combination of r spatial modes $\boldsymbol{\phi}_k(\boldsymbol{\xi})$ governed by time dynamics $a(t)$

$$\mathbf{x}(t) = \sum_{k=1}^{r} a_k(t)\boldsymbol{\phi}_k(\boldsymbol{\xi}). \tag{5.1}$$

When $r \ll n$, this is considered a low-rank embedding of the dynamics. For POD, the modes $\boldsymbol{\phi}_k(\boldsymbol{\xi})$ are determined from a singular value decomposition of a data matrix, and the dynamics are then projected into this space [98, 12]. DMD and mrDMD instead enforce that each mode be further constrained so that $a_k(t) = b_k \exp(i\omega_k t)$.

5.1.1 Dynamic mode decomposition

The DMD originated as a spectral decomposition for identifying coherent structures in fluid dynamics [137, 133]. Algorithmic refinements such as exact DMD [157] and practical developments [99] quickly followed, establishing DMD as a rigorous data-driven framework for spatiotemporal analysis and prediction. Many additional variants have been developed that capitalize on sparse,compressive measurements [85, 27], input-output systems and control [129], improved denoising [55, 77, 7], multiscale structure [100], randomized linear algebra [16, 63], and streaming data [78, 126].

The DMD separates spatiotemporal dynamics into a linear decomposition of coherent structures (DMD eigenmodes) with fixed temporal behavior (DMD eigenvalues). In particular, the DMD has connections to the Koopman operator which acts as the forward operator on scalar observables of a system. Given a discrete-time dynamical system

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k), \tag{5.2}$$

the Koopman operator $\mathcal{K}$ is defined as the linear forward operator acting on all scalar observables of the state [93, 94]

$$\mathcal{K}g(\mathbf{x}_k) = g(\mathbf{F}(\mathbf{x}_k)) = g(\mathbf{x}_{k+1}), \tag{5.3}$$

thus linearizing nonlinear dynamics to completely characterize the system. The approximation of Koopman eigenfunctions and eigenvalues through carefully selected observables is an active area of research which has given rise to Koopman mode decomposition [32], extended DMD [167], and kernel DMD [168]. DMD can be considered a special case of Koopman decomposition which restricts observables g to be point measurements of state. Thus the inputs to the DMD algorithm are the following data matrices of the measured state trajectory

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{m-1} \end{bmatrix} \\ \mathbf{X}' &= \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_3 & \dots & \mathbf{x}_m \end{bmatrix}, \end{aligned} \tag{5.4}$$

assumed locally related by a linear forward evolution map $\mathbf{A}$

$$\mathbf{X}' = \mathbf{A}\mathbf{X}. \tag{5.5}$$

DMD eigenvectors and eigenvalues are computed from a low-rank approximation of the forward evolution operator. The first step is computing the rank truncated POD modes of $\mathbf{X}$ via the SVD $\mathbf{X} = \mathbf{\Psi}_r \mathbf{\Sigma}_r \mathbf{V}_r^*$. Then the explicit expression for $\mathbf{A}$ is given by an expensive pseudoinverse operation

$$\mathbf{A} = \mathbf{X}' \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{\Psi}_r^*.$$

Instead of computing this large $n \times n$ matrix directly, we compute a low-rank projection onto the POD modes, $\tilde{\mathbf{A}} = \mathbf{\Psi}_r^* \mathbf{A} \mathbf{\Psi}_r$ which equals

$$\tilde{\mathbf{A}} = \mathbf{\Psi}_r^* \mathbf{X}' \mathbf{V}_r \mathbf{\Sigma}_r^{-1}.$$

The eigendecomposition of $\mathbf{A}$ can bow be efficiently computed from the eigenvalues and eigenvectors

$$\tilde{\mathbf{A}} \mathbf{W} = \mathbf{W} \mathbf{\Lambda},$$

where the diagonal entries of $\mathbf{\Lambda}$ are the DMD eigenvalues, and the DMD eigenmodes are computed from the scaled eigenvectors

$$\mathbf{\Phi} = \mathbf{X}' \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{W}. \tag{5.6}$$

This formulation by Tu et al [157] is known as exact DMD, and the authors prove that the DMD eigenmodes $\mathbf{\Phi}$ are the desired eigenvectors of $\mathbf{A}$.

Hence, DMD fits a linear approximation $\mathbf{A}$ to the system (5.2), but without explicitly computing $\mathbf{A}$. Rank truncation is achieved using the singular value decomposition to obtain $r \ll n$ eigenvectors and eigenvalues which evolve the system forward in time – a more efficient and interpretable representation of dynamics than $\mathbf{A}$ alone. Each state can be efficiently computed from a linear combination of DMD eigenvectors or eigenmodes (columns $\boldsymbol{\phi}_k$ of $\mathbf{\Phi}$), DMD eigenvalues ($\lambda_k = \mathbf{\Lambda}_{kk}$) and corresponding modal amplitudes $\mathbf{b} \in \mathbb{R}^k$

$$\mathbf{x}_k = \mathbf{\Phi} \mathbf{\Lambda}^k \mathbf{b}. \tag{5.7}$$

The equivalent expression in the continuous time setting uses a convenient scaling of the eigenvalues $\omega_k = \frac{\log \lambda_k}{i\Delta t}$

$$\mathbf{x}(t) = \sum_{k=1}^{r} b_k \boldsymbol{\phi}_k(\boldsymbol{\xi}) \exp(i\omega_k t). \tag{5.8}$$

Remark: The DMD approximation as stated introduces systemic bias in the eigenvalue computation of noisy data[55, 77]. To correct for this, we construct a weighted approximation of $\mathbf{A}$ *that incorporates both forward (*$\mathbf{X}' = \mathbf{A}_f\mathbf{X}$*) and backward (*$\mathbf{X} = \mathbf{A}_b\mathbf{X}'$*) time evolution, as formulated by Dawson et al [55]*

$$\mathbf{A} = (\mathbf{A}_f\mathbf{A}_b^{-1})^{1/2}. \tag{5.9}$$

This forward-backward DMD method is used throughout our results.

DMD provides a compelling description of many physical systems with low-rank, periodic temporal behavior, including fluid dynamics [158, 133], ocean sciences [100] and neuroscience [24]. The DMD approximation fails, however, when approximating intermittent and transient phenomenon, both of which violate the constrained temporal response in (5.8). In such cases, Fourier modes in time are insufficient to capture a richer set of dynamics.

5.1.2 Multiresolution DMD

The mrDMD method [100, 99, 101] circumvents some of the common shortcomings of standard DMD. Specifically, mrDMD provides a hierarchical temporal sampling framework, much like a wavelet decomposition, whereby disparate timescale phenomena can be characterized recursively. The recursive nature of mrDMD, illustrated in Fig. 5.1, can readily characterize distinct time scales by removing micro- and macro-timescale modes at different hierarchical levels. The mathematical structure of mrDMD accounts for the number of levels (L) of the decomposition, the number of time bins (J) for each level, and the number of modes retained at each level (m_L). Thus the solution is parametrized by the following three indices:

$$
\begin{aligned}
&\ell = 1, 2, \cdots, L, && \text{where } L = \#\text{ of decomposition levels} \\
&j = 1, 2, \cdots, J && \#\text{ time bins per level } (J = 2^{(\ell-1)}) \\
&k = 1, 2, \cdots, m_L && \#\text{ of modes extracted at level } L.
\end{aligned}
$$

To formally define the series solution for $\mathbf{x}_{\text{mrDMD}}(t)$, we propose the following indicator function

$$f_{\ell,j}(t) = \begin{cases} 1 & t \in [t_j, t_{j+1}] \\ 0 & \text{elsewhere} \end{cases} \quad \text{with } \ j = 1, 2, \cdots, 2^{(\ell-1)} \tag{5.10}$$

which is only non-zero in the interval, or time bin, associated with the value of j. The parameter ℓ denotes the level of the decomposition.

The three indices and indicator function (5.10) give the mrDMD solution expansion

$$\mathbf{x}_{\text{mrDMD}}(t) = \sum_{\ell=1}^{L} \sum_{j=1}^{J} \sum_{k=1}^{m_L} f_{\ell,j}(t) b_k^{(\ell,j)} \boldsymbol{\phi}_k^{(\ell,j)}(\boldsymbol{\xi}) \exp(i\omega_k^{(\ell,j)} t) \,. \tag{5.11}$$

This is a concise definition of the mrDMD solution that includes the information on the level, time bin location and number of modes extracted. Figure 5.1 demonstrates the mrDMD decomposition in terms of the solution (5.11). In particular, each mode is represented in its respective time bin and level. An alternative interpretation of this solution is that it generalizes the linear mapping (5.5) so that at each level ℓ of the decomposition one has

$$\mathbf{x}_{j+1}^{(\ell,j)} = \mathbf{A}^{(\ell,j)} \mathbf{x}_j^{(\ell,j)} \tag{5.12}$$

where the matrix $\mathbf{A}^{(\ell,j)}$ captures the dynamics in a given time bin j at level ℓ.

The indicator function $f_{\ell,j}(t)$ acts as a sifting function for each time bin. Interestingly, this function acts as the Gabór window of a windowed Fourier transform [98]. Since our sampling bin has a hard cut-off of the time series, it may introduce some artificial high-

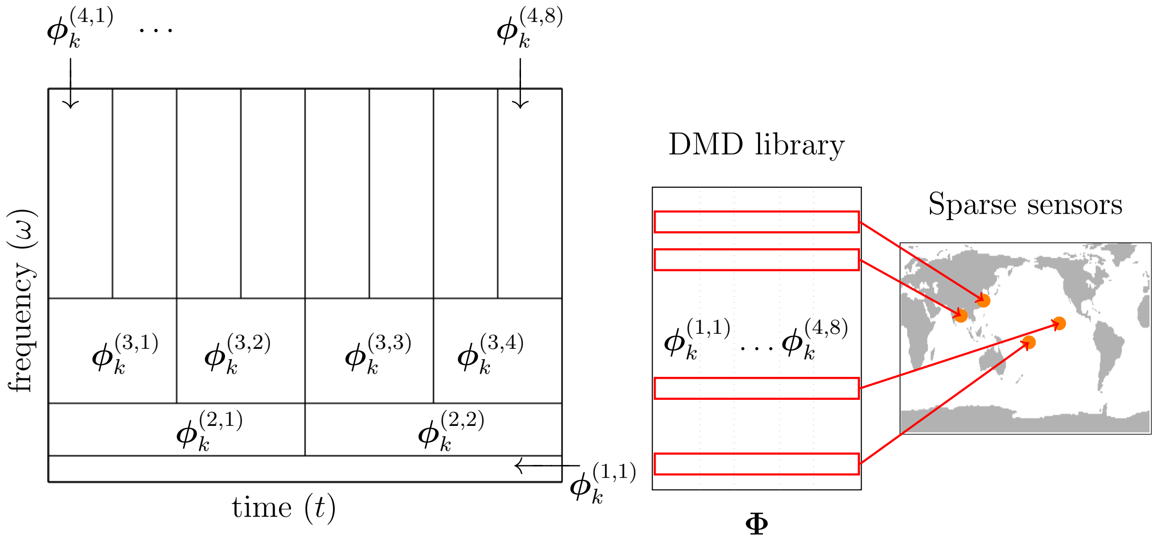


Figure 5.1: Illustration of the multiresolution analysis and sampling framework. On the left are the mrDMD modes $\boldsymbol{\phi}_k^{\ell,j}(\boldsymbol{\xi})$ and their position in the hierarchy. The triplet of integer values, ℓ, j and k, uniquely expresses the time level, bin and mode of the decomposition. Depicted in the middle is the matrix library of all mrDMD modes, which are then leveraged to obtain optimal samples (spatial sensors, right) for downstream estimation and prediction tasks.

frequency oscillations. Time-series analysis, and wavelets in particular, introduce various functional forms that can be used in an advantageous way. Thus thinking more broadly, one can imagine using wavelet functions for the sifting operation, allowing the time function $f_{\ell,j}(t)$ to take the form of one of the many potential wavelet bases, i.e. Haar, Daubechies, Mexican Hat, etc. Our approach uses a Haar-like windowing in time. A smoother temporal kernel can potentially identify randomly occurring dynamics, although this would be more computationally expensive.

5.2 Multiresolution analysis framework

We propose a multiresolution analysis decomposition and spatial sampling framework for the estimation and prediction of multiscale dynamics. Localized time-frequency analysis presents difficulties for future state prediction as the active modes in a new temporal window are unknown. Thus additional information is required to estimate temporal coefficients $\mathbf{a}(t)$. This additional information is commonly provided in the form of current state observations or sensor measurements. In this section we frame future state prediction as the reconstruction of new high-dimensional states from observations, or equivalently, estimating the correct temporal coefficients within the mrDMD library of modes. This is a common problem in engineering applications; the second half of this section describes how to optimize state measurement locations to minimize prediction error.

5.2.1 Mathematical formulation

The mrDMD yields a *library* or *dictionary* of possible dynamics which can be arranged in matrix form

$$\mathbf{\Phi} = \left[\boldsymbol{\phi}_k^{(1,1)} \quad \boldsymbol{\phi}_k^{(2,1)} \quad \boldsymbol{\phi}_k^{(2,2)} \quad \boldsymbol{\phi}_k^{(3,1)} \quad \ldots \quad \boldsymbol{\phi}_k^{(3,4)} \quad \boldsymbol{\phi}_k^{(4,1)} \quad \ldots \quad \boldsymbol{\phi}_k^{(4,8)}\right]. \tag{5.13}$$

There may be more than one mode per level indexed by k in the library $\mathbf{\Phi}$, which is now an $n \times M$ matrix. Thus any state can be approximated by a linear combination of the columns of $\mathbf{\Phi}$ with time-dependent coefficients

$$\mathbf{x}(t) = \mathbf{\Phi}\mathbf{a}(t). \tag{5.14}$$

The challenge with future state prediction is identifying the subset of $\mathbf{a}$'s components that are active (nonzero) at a given time, because mrDMD modes are localized in the time-frequency domain and do not enforce globally periodic temporal behavior.

The necessary information is often provided by point observations of the unknown state vector – an extremely prevalent construct in engineering and controls. Then there exists a linear relationship between observations, stored in a vector $\mathbf{y}$, and time coefficients. The components of $\mathbf{y}$ result from a linear observation map $\mathbf{C}$ applied to $\mathbf{x}$ and there is additive white noise $\boldsymbol{\zeta} \sim \mathcal{N}(0, \eta^2)$

$$\begin{aligned} \mathbf{y} &= \mathbf{C}\mathbf{x} + \boldsymbol{\zeta} \\ &= \mathbf{C}\boldsymbol{\Phi}\mathbf{a} + \boldsymbol{\zeta}. \end{aligned}$$

The following assumptions are made: $\mathbf{y} \in \mathbb{R}^p$, where $p \ll n$, and the measurement matrix $\mathbf{C} \in \mathbb{R}^{p \times n}$ consists of p rows of the $n \times n$ identity matrix which index measurements of the full state. This linear inverse problem for coefficient estimation is generally ill-posed (underdetermined) without additional regularization. We impose two forms of regularization:

1. **Coefficient sparsity.** To reflect the physical constraint that only a few modes are active at a given time, we impose a sparse structure on $\mathbf{a}$ which minimizes the number of nonzero entries (see below).

2. **Designing C.** Point observations of states will be optimized to leverage mrDMD features of interest and eliminate redundancy between point samples. In engineering settings this is known as sensor placement.

At first glance both pose combinatorial subset selection problems, but as we shall see, there exist well-known methods to approximate both selections.

5.2.2 Online estimation and prediction

Assume a fixed $\mathbf{C}$ that is already optimized for estimation. The sparsest coefficients $\mathbf{a}_\star$ can be approximated using a convex l_1 constrained minimization

$$\mathbf{a}_\star = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_1 \text{ subject to } \mathbf{y} = \mathbf{C}\boldsymbol{\Phi}\mathbf{a}. \tag{5.15}$$

This search for the sparsest coefficients in a generic dictionary is known as *basis pursuit* [154]. The specific application of recovering the sparsest coefficients in a *tailored* dictionary of learned features is known as sparse representation for classification or SRC [170]. The sparsest solution $\mathbf{a}_\star$ can be recovered given it is sufficiently sparse in this dictionary and given sufficiently many measurements. The required number of measurements depends on r, the number of nonzero coefficients in $\mathbf{a}_\star$. A sufficient (not necessary) condition for the number of random measurements for recovery using a generic dictionary is given by $p \sim O(r \log(n/r))$ [58, 35, 8]. In practice, SRC recovers the solution for tailored dictionaries using drastically fewer measurements [170, 106, 21].

We further assume the noise distribution is unknown in which case it is appropriate to solve the noiseless basis pursuit problem with an equality constraint. This is acceptable for the small levels of observation noise in our examples and is shown to be robust to noise in various applications [170, 106]. For high levels of noise one can solve a relaxed basis pursuit problem in which the least-squares error is a parameter derived from an estimate of the noise level.

The basis pursuit optimization(5.15) is convex, and thus can be efficiently solved using greedy algorithms available in l_1 optimization packages (CVX [73], SPGl1 [160] and CoSaMP [118]). We use a spectral gradient-projection method [161] implemented in SPGl1, which attempts to balance the tradeoff between least-squares fit and solution sparsity.

The full state is subsequently estimated from the active (nonzero) modal dynamics using $\hat{\mathbf{x}} \approx \mathbf{\Phi}\mathbf{a}_\star$ (see (5.14)). A more accurate reconstruction can be achieved with traditional least-squares approximation using only the active modes. Explicitly, we first determine which library modes are active by indexing the nonzero solution coefficients

$$\boldsymbol{\gamma} := \{i \in \{1, \ldots, r\} \mid a_{\star_i} \neq 0\}. \tag{5.16}$$

Once $\boldsymbol{\gamma}$ is known, then state reconstruction reduces to a least-squares approximation using the subset of active modes within the library

$$\begin{aligned} \hat{\mathbf{x}} &\approx \mathbf{\Phi}_\gamma \mathbf{a}_\gamma \\ &= \mathbf{\Phi}_\gamma (\mathbf{C}\mathbf{\Phi}_\gamma)^\dagger \mathbf{y}. \end{aligned}$$

When $\mathbf{\Phi}$ are POD modes, this is widely known as gappy POD, introduced in the early 90s for image reconstruction using eigenfaces [65] and subsequently used for flow field reconstruction in ROMs [165]. Sparse approximation methods similar to the above description have applied compressed random measurements in POD mode libraries to classify different parameter regimes [29, 21]. In recent work [96] classification is accomplished with sparse measurements in a DMD mode library. However, to our best knowledge, this is the first known application of sparse estimation using a DMD mode library samples with optimized measurements.

5.2.3 Offline measurement selection

We seek to the optimal design of the measurement operator to obtain a minimal set of non-redundant observations of the state. The repetition of modes across multiple time-frequency bins in the multiscale library introduces redundancy. Optimal measurement design from this library would then result in multiple measurement locations with the same dynamics, which is undesirable in practice. This is remedied by first preprocessing the library and constructing an index set to filter out redundant modes (this is distinct from the local active set $\boldsymbol{\gamma}$). For example, this subset can be selected from modes with dominant spatiotemporal signatures, by identifying mrDMD modal amplitudes which exceed some threshold T

$$\boldsymbol{\alpha} := \{i \in \{1, \dots, m\} \mid |b_i| > T\}. \tag{5.17}$$

Alternatively, depending on the application, modes corresponding to specific time or frequency bins of the decomposition can be extracted in this step for downstream sensor selection. The mathematical aim is to construct the measurements $\mathbf{C}$ to minimize the least-squares approximation error, $\mathbf{a} - \hat{\mathbf{a}}$. The number of point measurements, p, has to be greater than or equal to m, the number of columns of $\mathbf{\Phi}_{\boldsymbol{\alpha}}$, which is dictated by the dynamics of interest. Because no more than m modes are active in the training window, this guarantees that chosen sensors are designed to capture all possible active dynamics. We prescribe the smallest possible choice of p (to be equal to m), although oversampling is also possible when

$p > m$.

Recall the the column space of the multiscale library spans the spatial variable, and as before, row selection of mrDMD modes using QR pivoting serves to maximize the dynamical information carried by the modes. Unlike other sensor placement methods, QR pivoting can take any matrix as input, and easily generalizes to basis functions other than POD (PCA). For this step the library $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$, number of measurements p and noise distribution are assumed predetermined, and point measurements are selected to minimize some scalar measure of the "size" of the error covariance

$$\boldsymbol{\Xi} = \mathrm{Var}(\mathbf{a} - \hat{\mathbf{a}}) = \sigma^2[(\mathbf{C}\boldsymbol{\Phi}_{\boldsymbol{\alpha}})^T \mathbf{C}\boldsymbol{\Phi}_{\boldsymbol{\alpha}}]^{-1}. \tag{5.18}$$

Note that this covariance matrix is analogous to that obtain from POD modes in Chapters 3 and 4. As before, we compute p optimal point measurement indices from the QR factorization pivots of the dominant dynamical modes $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$

$$\boldsymbol{\Phi}_{\boldsymbol{\alpha}}^T \mathbf{C}^T = \mathbf{Q}\mathbf{R}, \tag{5.19}$$

which is described in Algorithm 2.

5.3 Applications

This section demonstrates the application of our multiresolution analysis and sampling methods on data presenting multiscale temporal scales. The first example, a manufactured video, presents three spatial modes independently oscillating at different frequencies in overlapping time windows. We empirically quantify the accuracy of our analysis and estimation from optimized samples by comparing with the known dynamics. By contrast, the second example seeks to identify intermittent warming phenomena from real global satellite data of sea surface temperatures. Here the accuracy metric is the correct identification of warming events in the validation window based on the same events identified in a separate training window. For both examples, results from the multiresolution analysis are contextualized with appropriate comparisons to POD and DMD.

5.3.1 Multiscale video example

The video data is generated from three spatial modes oscillating at different frequencies in overlapping time intervals, effectively appearing mixed in time. Mathematically each snapshot is constructed as a linear combination of Gaussian spatial modes

$$\mathbf{x}(t) = \sum_{i=1}^{3} a_i(t) \exp\left(-\frac{(\boldsymbol{\xi} - \boldsymbol{\xi}_{0_i})^2}{w_i} \right), \tag{5.20}$$

where $\boldsymbol{\xi} \in \mathbb{R}^n$ is the vectorized planar spatial grid ($n = nx \times ny$) and w is a scalar width parameter. The challenge with mrDMD is to identify the true modes $\boldsymbol{\phi}_i = \exp\left(-\frac{(\boldsymbol{\xi} - \boldsymbol{\xi}_{0_i})^2}{w_i} \right)$ within their respective temporal windows. We list explicitly the modal components with their time dynamics in table 5.1 and fig. 5.2.

Table 5.1: Multiscale video dynamics

Mode	Time dynamics	frequency (ω_i)
$\boldsymbol{\phi}_1$	$a_1(t) = \begin{cases} \exp(2\pi i \omega_1 t) & t \in [0, 5] \\ 0 & \text{elsewhere} \end{cases}$	$\omega_1 = 5.55$
$\boldsymbol{\phi}_2$	$a_2(t) = \begin{cases} \exp(2\pi i \omega_2 t) & t \in [2.5, 7.5] \\ 0 & \text{elsewhere} \end{cases}$	$\omega_2 = 0.9$
$\boldsymbol{\phi}_3$	$a_3(t) = \begin{cases} \exp(2\pi i \omega_3 t) & t \in [0, 10] \\ 0 & \text{elsewhere} \end{cases}$	$\omega_3 = 0.15$

The identification of spatial modes clearly depends on separation by temporal frequency. fig. 5.3 compares mrDMD results with proper orthogonal decomposition (POD), the primary dimensionality reduction tool used in ROMs. The mrDMD time-frequency analysis successfully isolates the true modes of the system, while POD fails at this task since it is a variance-based decomposition in which modes are eigenvectors of data covariances. Hence,

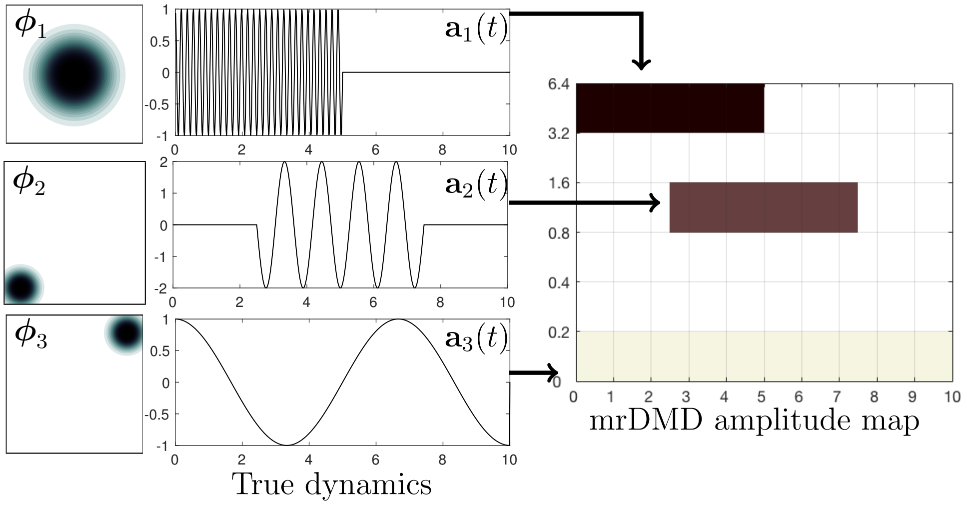
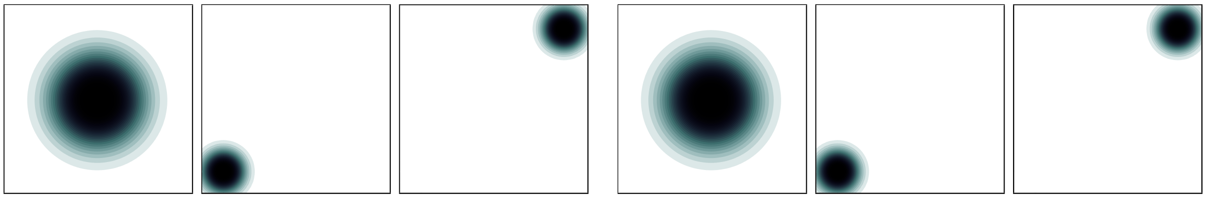


Figure 5.2: **Multiscale video**. The mrDMD analysis accurately identifies the active frequencies in each temporal window of the decomposition. The plot on the right colors each time-frequency bin by mrDMD mode amplitude.

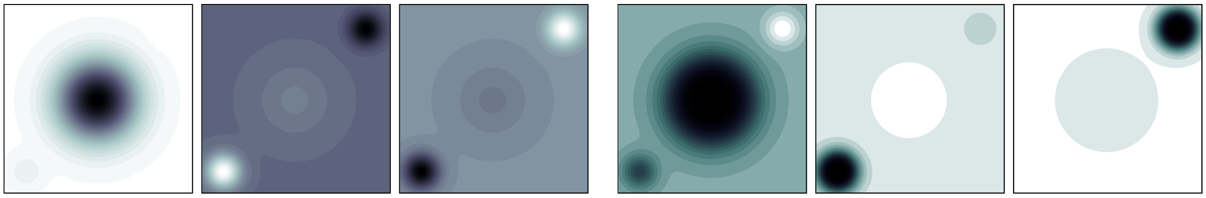
the POD spatial modes appear mixed. In addition, mrDMD outputs the correct frequencies associated with each of these modes. We do not give results for standard DMD since it will fail by fitting exponentials across the global time window. Within the multiresolution analysis, active modes are repeated across the time-frequency bins, so we include only the mode with highest amplitude. These selected modes describe the feature selection or formation of the index set $\boldsymbol{\alpha}$ used to train sensors. Similarly, we select the first three POD modes, which explain 99.9% of the variance in the training data, as quantified by the POD eigenvalues. Interestingly, both decompositions yield the same optimal samples for this example. This is because the three point sources of diffusion correspond to spatial extrema, which are indicated by the overlapping contour plots of the first three true, mrDMD and POD modes respectively.

Even though the same optimal samples are identified, feature selection greatly affects the quality of reconstruction for future state prediction. To test this, we generate test data using time coefficients for each mode from a random process in order to estimate the true



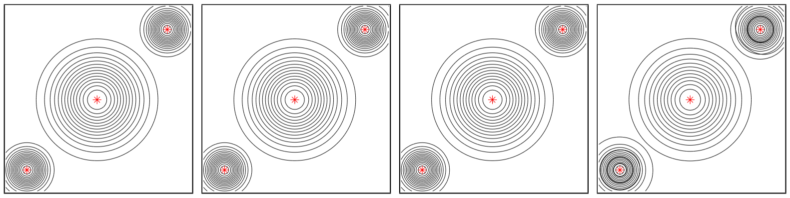
(a) True modes 1-3

(b) mrDMD modes 1-3



(c) DMD modes 1-3

(d) POD modes 1-3



(e) QR sensors, left to right: true, mrDMD, DMD, POD

Figure 5.3: **Comparison of spatial modes and sensors** shows that mrDMD (b) is ideal for isolating the true windowed dynamics (a), unlike standard DMD (c) or POD (d). Despite varying degrees of spatial mode separation, QR sensors extracted from the different basis modes (e) closely agree.

coefficients from the three optimal sampling locations. Further, white noise of variance $\sigma^2 = 0.001$ is added to the three sampled observations. Finally, coefficients estimated from the subselected basis $\mathbf{\Phi}_{\boldsymbol{\alpha}}$ consisting of mrDMD modes (fig. 5.3b) are compared to a basis of POD modes (fig. 5.3d). Here least-squares estimation is used since the problem is well-posed (three equations and three unknowns). Results are plotted in fig. 5.4, where it can be seen that mrDMD-based estimation (blue) best approximates the true random process (black). By contrast, POD-based estimation (red) deviates from the true coefficients, particularly in the second and third modal coefficients.

We quantify this further in a noise study whose results are plotted in fig. 5.5. White

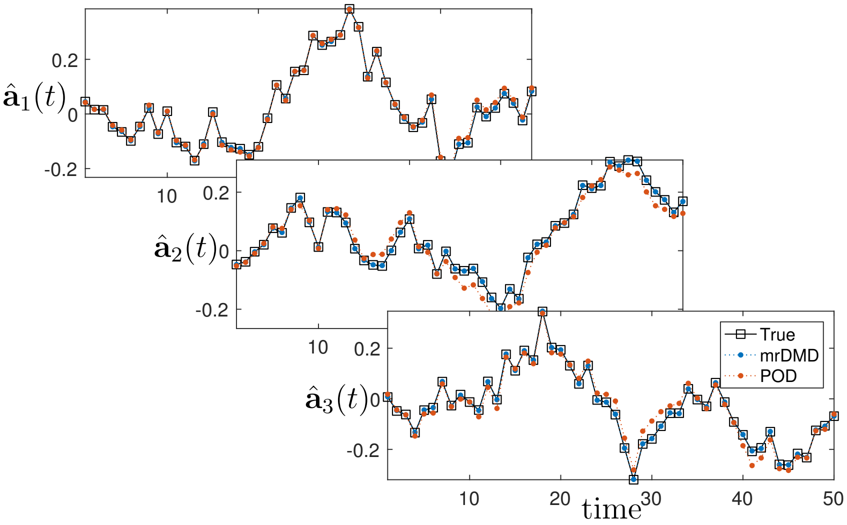


Figure 5.4: **Comparison of estimated coefficients**. The estimation of time coefficients from mrDMD modes and QR-optimized sensors is more accurate than with POD modes. True coefficients in this test window are generated by a Gaussian random process.

noise levels σ^2 are increased on a logarithmic scale to examine the l_2 reconstruction error trend of the coefficient vectors. Estimation with mrDMD points to an exponential increase in error until the signal appears saturated by noise at $\sigma^2 = 1$. POD-based estimation, however, appears to saturate much earlier at small noise levels, and is thus not as suitable for multiscale estimation and prediction. Specifically, the full state can now be reconstructed from noisy point observations in any validation window using estimated coefficients eq. (5.14).

In this video example, the true spatiotemporal modes were selected based on three different identified frequencies. For estimation, we leverage these three modes and associated sensors to reconstruct unseen random dynamics at future time instances. In many systems, it is not immediately obvious which of the identified dynamics are active in a new time window. This lack of knowledge of true dynamics typifies many real-world systems, in which nonlinear, chaotic dynamics and regime switching limit our ability to predict future states. However, we can harness observations from optimally sampled observations (sensors) to classify active dynamics in new time windows, which is illustrated here for sparse identification

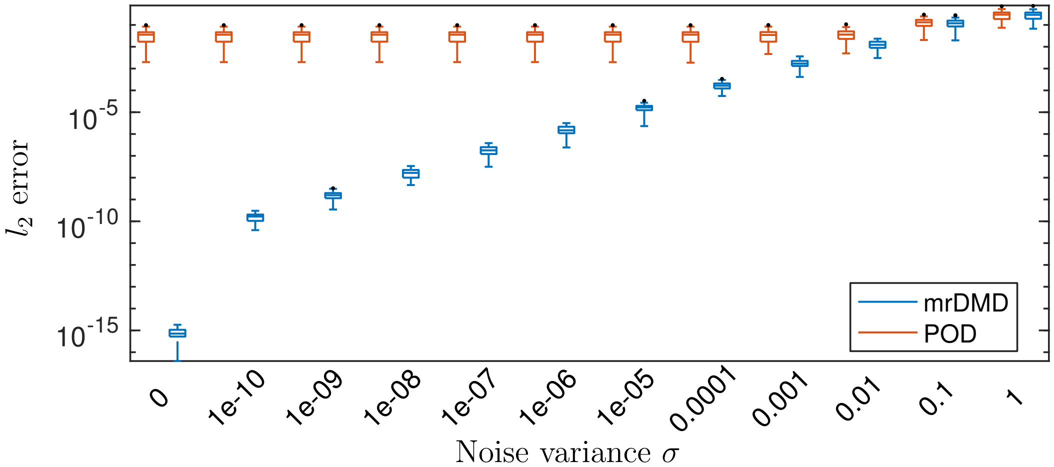


Figure 5.5: **Noise corrupted measurements**. Approximation error with mrDMD modes and sensors trends several orders of magnitude lower than with POD modes and sensors for small levels of sensor white noise.

of El Niño warming periods from global ocean temperature data.

5.3.2 NOAA ocean surface temperature

We apply multiresolution analysis to real satellite data of weekly ocean surface temperatures from 1990-2017. Here, the measure of success is accurate identification of intermittent warming and cooling events that are famously implicated in global weather patterns and climate change - El Niño and La Niña events. Weekly temperature means are collected on the entire 360x180 spatial grid, with data omitted over continents and land. The dimension of each weekly snapshot then reduces from $360 \times 180 = 64800$ to $n = 44219$ spatial gridpoints. We run mrDMD in training windows of 16 years (multiples of 2 facilitate clean separation of annual scales) up to four decomposition levels so that the fastest identifiable frequency is biennial, effectively discarding the dominant annual scale from the analysis. This is done to parse out the El Niño Southern Oscillation (ENSO), defined as any sustained temperature anomaly above running mean temperature with a duration of 9 to 24 months.

Figure 5.6 visualizes the resulting modal amplitude maps in the time-frequency domain, along with oscillatory dynamical modes identified by the decomposition. Each bin is colored

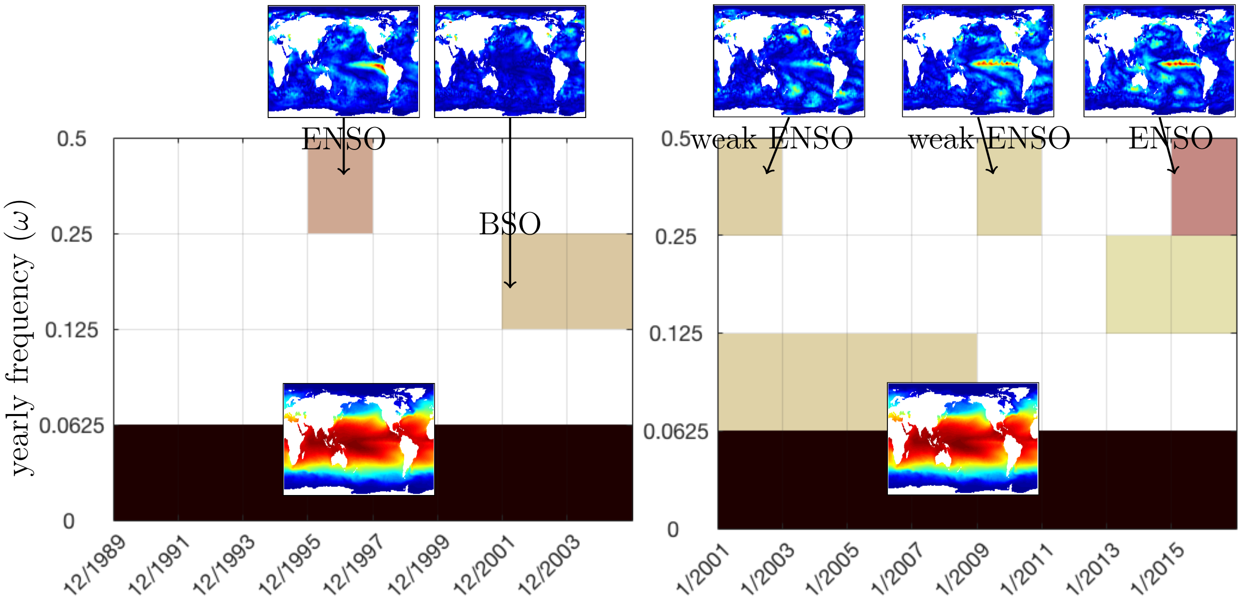
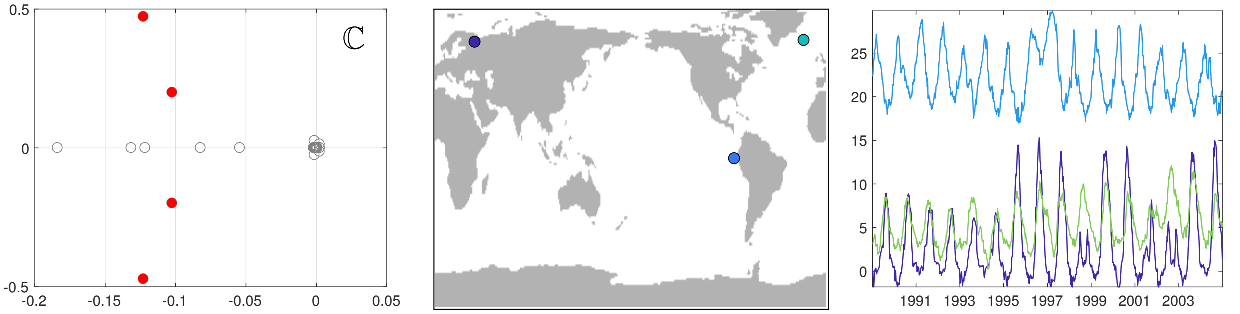


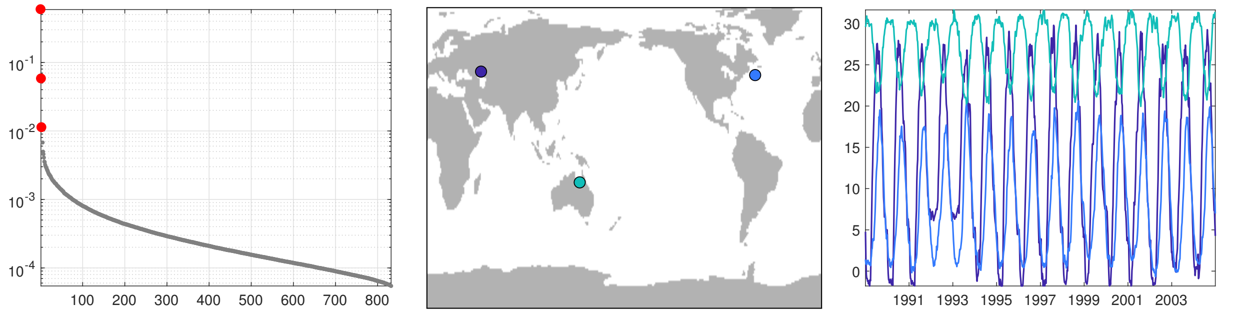
Figure 5.6: **MrDMD map of modal amplitudes by level and time window**. The method captures several dynamically significant oscillations occurring slower than annual dynamics. The identified El Niño (ENSO) and Barents Sea oscillations (BSO) are both rare events known to exert significant influence on global weather patterns.

by the average modal amplitude of identified dynamics. An energetic background mode is identified in both training windows (1990-2006, 2001-2016) corresponding to the mean temperature distribution. More importantly, lower-energy dynamics identified at the biennial scale correspond closely to well-documented ENSO events, the strongest of which occur in 1997-1999 and 2014-2016. Visual inspection clearly identifies these as ENSO events which cause a characteristic band of warm water across the South Pacific near coastal Peru.

We study now the consequences of multiscale analysis on optimal sampling, which can be readily interpreted as physical sensors in this application. Figure 5.7 compares mrDMD and POD in terms of spectral and spatiotemporal information relayed by mrDMD and POD-based sensors. DMD eigenvalues with nonzero imaginary part (blue, fig. 5.7a) identify active oscillations within the 1990-2006 training window. One of the conjugate frequency pairs confirms the true ENSO yearly frequency with $\text{Im}(\omega) = 0.5$. On the other hand, the POD



(a) mrDMD frequencies ω_k (b) mrDMD sensors from QR (c) Sensor time series



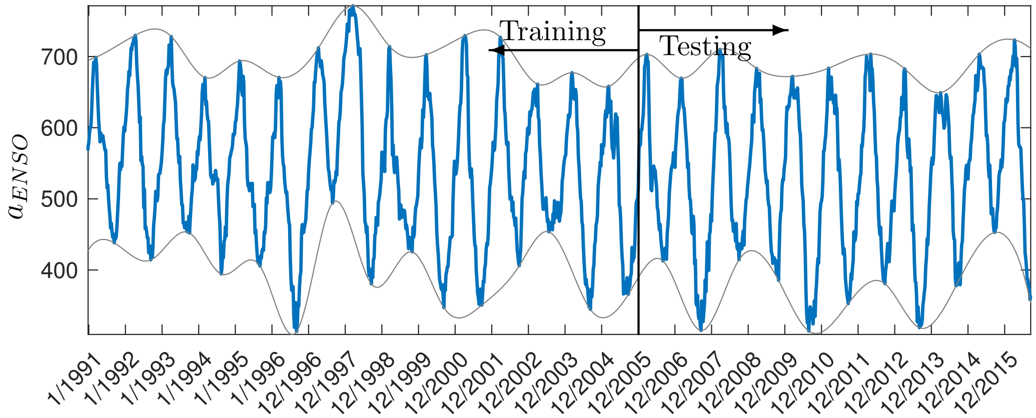
(d) Singular values (e) POD sensors from QR (f) Sensor time series

Figure 5.7: **mrDMD vs. POD sensors.** QR sensors sampled from the ENSO and BSO mrDMD modes (see fig. 5.6) identify the relevant warming region in the South Pacific, unlike the POD sensors. The time series from mrDMD sensors also display richer temporal trends compared to the dominant annual trends seen in the POD time series.

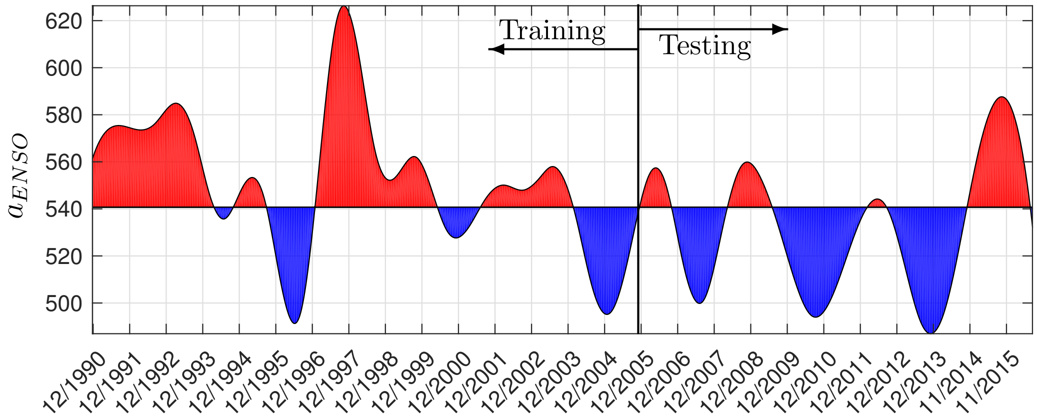
spectrum characterizes data covariances and exhibits slow decay due to noisy, low-energy dynamics in the system. This prohibits a low-rank representation of this data using a few POD modes. We now compute sensors from both sets of modes. We form $\mathbf{\Phi}_{\boldsymbol{\alpha}}$ from the oscillatory dynamics in the mrDMD spectrum, and the first three POD eigenmodes in the POD spectrum. QR-based sensors are computed for both sets of modes separately. The first observation is that, unlike the toy video example, mrDMD and POD sensors are different. One mrDMD sensor is located over the ENSO band, proving that multiscale characterization is crucial for this system. Meanwhile, POD sensors appear sensitive to temperature extremes occurring along coastlines and inland seas. POD sensor measurements also entirely miss the ENSO phenomenon in the South Pacific. This can be seen from the time series measured at all sensor locations – mrDMD time series reflect a richer set of dynamics than those of POD, in which only a dominant annual scale is present. The comparison with POD-based sampling is particularly valuable and demonstrates POD's inability to isolate low-energy intermittent ENSOs in the training window. POD is unable to obtain a low-rank representation of the dynamics because the ocean temperature system is driven by many more degrees of freedom than the toy example. By contrast, mrDMD curtails artificial rank inflation through localized time-frequency analysis.

Optimal sampling can be exploited in a data assimilation-type framework in which prediction is accomplished by parameter estimation followed by full state reconstruction from sensors. We estimate ENSO intensity in a validation window using sparse estimation of library ($\mathbf{\Phi}$) coefficients eq. (5.15). Figure 5.8b plots the signal envelope in both the training and validation windows, then traces temperature anomalies about a baseline using the mean of upper and lower envelopes. The resulting measure, which demarcates events above and below baseline (red and blue respectively), closely agrees with well-documented El Niño and La Niña events.

Full state snapshots of ocean temperatures may be reconstructed using the identified coefficients as described in section 5.2.2. In addition, our sampling framework can be used as a form of multiscale model reduction. Traditionally, ROMs include a POD Galerkin projection to reduce dimensionality of the original system. However, in general DMD modes are not orthonormal so the projection cannot be easily inverted to lift back to the full state. We show here how a few active mrDMD modes can be used for global reconstructions of the state from 30 sensors in the temperature field. Section 5.3.2 displays the number of times a spatial location is selected as a QR pivot in each temporal window of the decomposition. As expected, spatial locations corresponding to the ENSO mode are selected with less frequency, reflecting the intermittency of these dynamics. Finally, fig. 5.10 displays the results of state reconstruction at different time windows using optimal samples.



(a) Envelope of ENSO coefficient from training window 1990-2006



(b) Prediction envelope mean

Figure 5.8: **ENSO coefficient prediction**. Coefficient estimation of ENSO mode from only 2 sensors accurately predicts El Niño phenomena. Red peaks closely agree with well-documented strong ENSO events in 1997-99, 2014-2016, as well as with weak ENSO events in 2002-3, 2006-7 and 2009-10. Cooling events (blue) agree with documented La Niña events in 2007-8, 2008-9 and 2010. Small discrepancies in the identified times can be explained by natural variations between the various warming events.

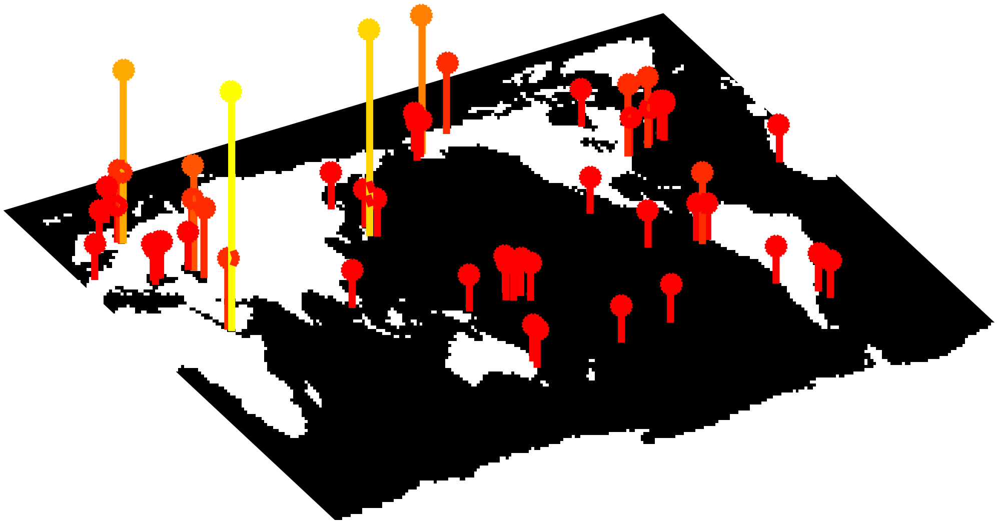
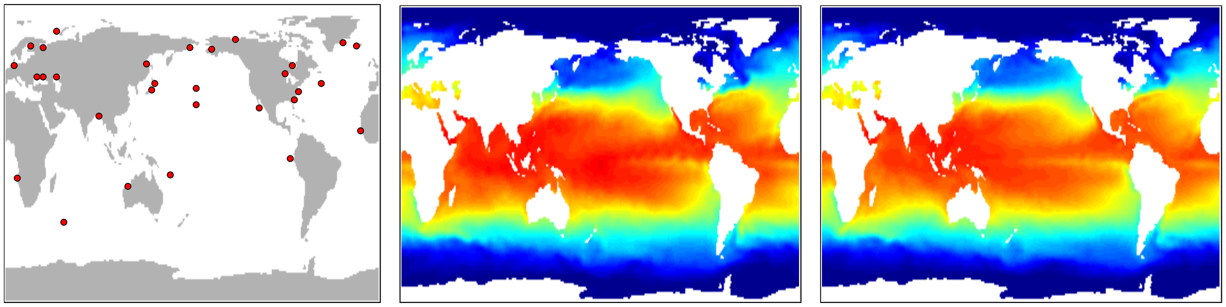


Figure 5.9: **Sensor Ensemble**. Time-dependent sensor aggregation pinpoints low-energy locations (i.e. ENSO sensor along coastal Peru) that would have otherwise been ignored by a global analysis.

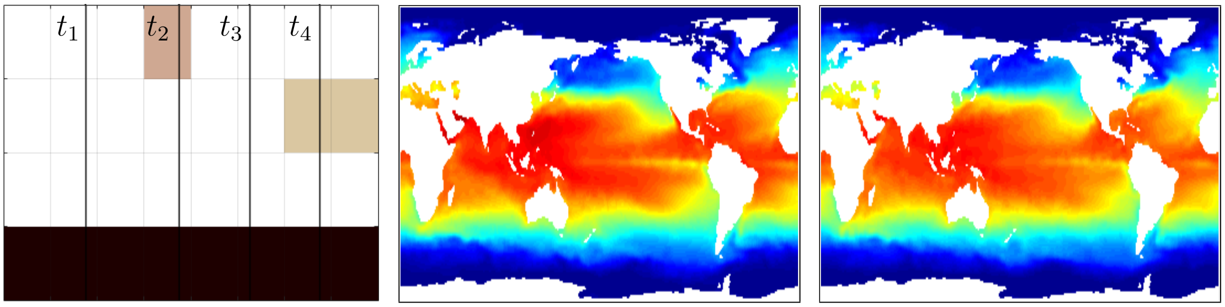
5.4 Conclusions

Despite significant computational and algorithmic advances, the modeling and prediction of multiscale phenomena remains exceptionally challenging. Much of the difficulty arises from our inability to disambiguate the underlying and diverse set of governing dynamics at various spatiotemporal scales. Additionally, for large-scale systems, our limited number of sensor measurements can greatly restrict model discovery efforts, thus compromising our ability to produce accurate predictions. In this chapter, we have shown that the multiscale disambiguation and limited measurement problem can be simultaneously addressed with principled, algorithmic methods. Specifically, combining the multiresolution dynamic mode decomposition with QR column pivots gives accurate spatiotemporal reconstructions and predictions for intermittent multiscale phenomena.

In broader context, our algorithmic developments provide a principled framework for understanding the optimal placement of sensors in complex spatial environments and/or networked configurations. The greedy QR column pivot selection used for selecting sensor



(a) mrDMD sensors from QR (b) t_1=6/1992 (c) t_2=6/1996



(d) mrDMD map 1990-2006 (e) t_3=6/1999 (f) t_4=6/2002

Figure 5.10: **Reconstruction from 30 sensors in flow field.** MrDMD reconstruction from sensors is reasonably accurate across snapshots selected from different time windows of the decomposition. Each snapshot of 44219 gridpoints is approximated to less than 5% relative error.

locations leverages the dominant low-rank features of the multiscale physics in order to best predict and reconstruct the spatiotemporal dynamics. By incorporating a multiresolution analysis tool, i.e. the mrDMD, respect is also given to the diverse temporal phenomena that are often observed in practice. By systematically accounting for multiscale physics, this placement of sensors performs significantly better than the same limited number of sensors that do not account for these physics.

The optimized sampling and prediction algorithms developed here have potential for a wide variety of technological applications. Indeed, with the global increase in sensor networks and sentinel sites for monitoring, for instance, ocean and atmospheric dynamics, disease spread across countries, and/or chemical pollutants, new methods are needed that respect both limited budgets (i.e. limited sensors & cost) and multiscale physics. To our knowledge, our algorithms are the first to simultaneously take into account both of these critical components in an efficient, scalable algorithm.

Chapter 6

SPARSE APPROXIMATION FOR INSECT FLIGHT DYNAMICS

Closely related to the concept of reconstruction is classification, where instead of recovering the full state from measurements, we classify measurements into certain parameter regimes of the attractor. Intuitively classification should require fewer measurements than reconstruction using gappy POD or compressed sensing. This principle is reflected in many natural phenomena, whereby insects, animals and robots interacting with high-dimensional flow fields need not internally reconstruct them to make critical prediction and control decisions. In fact, they are able to do so with extremely sparsely distributed, noisy measurements of strain or pressure.

In this chapter, we present a hybridized approach that leverages convex optimization and optimized measurements to *classify* dynamics within a *data-driven library* of parametrized flow regimes. This method is centered around an application on insect wing strain measurements in flight which are generated by a flexural wing chord forced by different frequencies of dynamical regimes.

6.1 Motivation

There is evidence that wings harness low-dimensional flow structure such as leading edge vortices to maximize lift and improve stability in flight [15, 163, 53]. This is consistent with the observation that wing motion is itself low-dimensional. Proper orthogonal decomposition analysis has revealed that fin motion [147], bat [132] and avian wing motion are constrained to only a few degrees of freedom. The lack of complex musculature in insect wings results in even fewer degrees of freedom. Such dimensionality reduction suggests that insects, which are constrained to small ranges of motion, capitalize on low-dimensional feature spaces to inform low-dimensional control protocols.

Wing mechanosensors in insects detect and leverage these fluid forces and accelerations for changing flight environments. Indeed, tactile sensory mechanisms have been identified on the wings of most animals - bats sense with small hairs on their wings [140], birds sense with wing feathers [22], but insects possess a small number of strain sensors on their wings called *campaniform sensilla*. The *sensilla* are strongly implicated in neurosensory flight control [57, 135], in part because insect wings react to disturbances faster than visual stimulus transmission to the central nervous system [49]. Aerodynamic feedback is encoded within the strain signals, registering fluidic loading frequencies into the signal that can be exploited for characterizing flow environment. Evidence shows that insects exploit innate or learned knowledge of fluid environment through the strain encodings to make split-second decisions in flight. The insect nervous system has evolved specifically for the decision task, but controllers in hovercraft and MAVs require other means of exploiting point sensor feedback. Strain sensors are too sparsely distributed to fully resolve spatial flow encoding over the wing, and equation-based flow identification or prediction is expensive and difficult to generalize to different flow regimes.

The focus of this chapter is sensor-enabled, data-driven flow environment classification using sparse approximation of incoming sensor measurements. Our data consists of strain point measurements from a numerical fluid-structure interaction model of a hawkmoth wing undergoing different environments of flow feedback. First, POD modes of strain measurements from different aerodynamic environments are assembled into a dictionary of low-rank dynamical states labelled by environment. This supervised learning stage mimics experiential learning and trains the data-driven model specifically for the task at hand. Next, incoming strain measurements are classified in this learned POD library using ℓ_1 constrained sparse approximation and sparse representation for classification (SRC) [170]. Importantly, sparse approximation aligns with plausible neurobiological sparse encoding strategies [122].

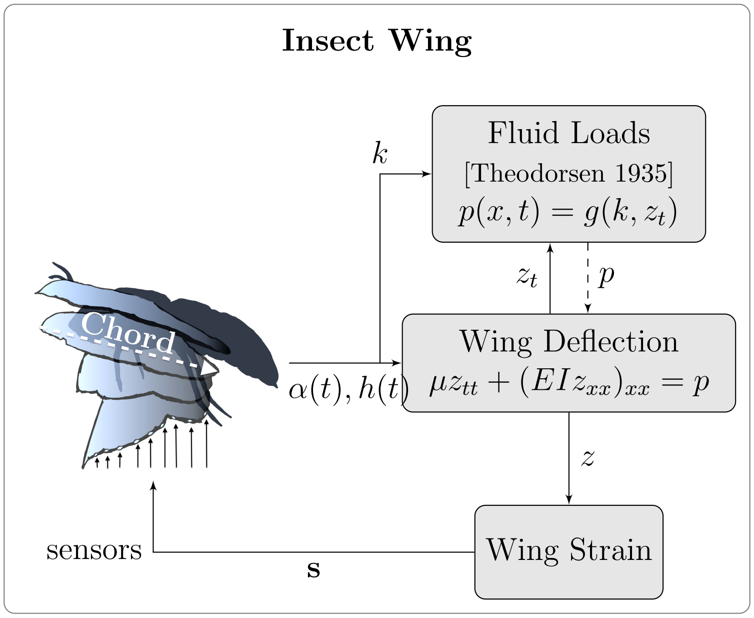


Figure 6.1: Strain sensors in insect wings directly encode elastic wing deformation from the leading edge dynamics and flow feedback loads. These fluid loads are driven by dominant flapping and external reduced frequencies k and weak coupling from wing deformation.

6.2 Numerical Model of Fluid-Structure Interaction

We describe our numerical approach for analyzing the fluid-structure coupling of the wing, and motivate sparse representation for classification in the frequency domain (SRCf). The wing is modeled by classical elasticity theory [74]. The governing equations model the deformation of the elastic body subject to a function of the time-dependent chord loading $p(x,t)$. We define the spatial domain x to represent the length of the wing chord that deforms in response to inertial-elastic forces. The range of x is the closed interval $[0, 2b]$ where b is the half-chord length, $x = 0$ represents the *leading edge* undergoing oscillatory actuation, and $x = 2b$ is the stress and shear free *trailing edge.* These boundaries are consistent with biology - most insect wings are actuated by exoskeletal structure at their leading edge, while their trailing edge consisting primarily of thin membrane flutter at the behest of air loads without experiencing any inertial stress or shear forces. True wings deform in three dimensions,

Nomenclature	
b	Half-chord length of wing [.01 m]
c	Categories or number of environments
$C(k)$	Theodorsen transfer function
$EI(x)$	Flexural stiffness of wing [Nm^2]
f	Frequency of maneuver [Hz]
$\mathbf{f}$	Vector of frequencies
k	Reduced frequency [$k = 2\pi fb/U_\infty$]
h_0	Initial vertical position of wing
$h(t)$	Vertical position of plate
m	Spatial grid resolution
N	Number of timesteps
$p(x,t)$	Chordwise loading of fluid pressure
$\mathbf{s}$	Vector of transverse strain of wing
$\mathbf{S}$	Matrix of transverse strains
$\hat{\mathbf{S}}$	Discrete Cosine Transform of $\mathbf{S}$
t	Time [s]
$\mathbf{t}$	Vector of time t in seconds
U_∞	Free stream velocity [10 m/s]
x	Chordwise spatial coordinate [m]
$z(x,t)$	Transverse deflection of wing [m]
α_0	Base angle of attack
$\alpha(t)$	Angle of attack of wing
ω	Angular velocity of maneuver
μ	Linear wing density [.002 kg/m]

Table 6.1: This table presents some additional notation used in this chapter.

however we can approximate the deforming wing using two spatial dimensions x and z where $z(x)$ is the vertical displacement that depends on position along the chord. Assuming normal deflections and ignoring rotational effects, the elastic wing can be characterized by a forced linear Euler Bernoulli beam with spatially varying flexural stiffness $EI(x)$ and a zero deformation initial state that corresponds to a flat stationary wing. However, the strain data is collected after the deformation achieves a steady state, ignoring any initial transients. The governing equations and boundary conditions are given by:

$$\mu z_{tt} + (EIz_{xx})_{xx} = p(x,t), \qquad x \in (0, 2b), t \geq 0 \tag{6.1a}$$

$$z(x,0) = 0 \qquad z_t(x,0) = 0 \tag{6.1b}$$

$$z(0,t) = h(t) \qquad z_{xx}(2b,t) = 0 \tag{6.1c}$$

$$z_x(0,t) = -\sin(\alpha(t)) \qquad (EIz_{xx}(2b,t))_x = 0 \tag{6.1d}$$

where $z(x,t)$ is the displacement of the elastic body along the length $x \in (0, 2b)$, $\alpha(t)$ and $h(t)$ specify the leading edge time dynamics for pitching and plunging respectively, and $p(x,t)$ is the applied pressure loading that is computed from the Theodorsen model. The trailing edge boundary conditions specified at $x = 2b$ model a shear-free and stress-free boundary.

Numerical simulations use wing morphological parameters of the *Manduca sexta* hawkmoth, a representative species capable of hovering and other small-scale maneuvers that typify insect flight. The numerical model consists of two components that are weakly coupled as shown in Figure 6.1: (1) elastic deformation strains and (2) unsteady aerodynamic loads. The deformation $z(x,t)$ is simulated with the hawkmoth pitching and plunging motion at 26 Hz at the leading edge. The *Manduca* wing's morphological parameters [52, 51] facilitate analysis of a small parameter space of a simplified 2D domain - the deforming wing chord along the xz plane. Accordingly, the flexural stiffness distribution $EI(x) = 10^{-5}e^{-150x}$ is determined by averaged experimental values from the actual hawkmoth in [52], in which wings are shown to exhibit exponentially decreasing chordwise wing stiffness away from the leading edge. An implicit finite-difference scheme is used to solve the dynamic 1D beam equation for the normal deformation $z(x,t)$ with spatially varying flexural stiffness to investigate the resulting unsteady fluid pressures and lift forces. Denoting the vector of transverse deflection at time t_i as $\mathbf{z}_{t_i}$, the strain vector $\mathbf{s}$ is computed from the componentwise operation on $\mathbf{z}$,

$$\mathbf{s}(\mathbf{x},t_i) = \frac{\sqrt{\Delta\mathbf{z}_{t_i}^2 + \Delta x^2} - \sqrt{\Delta\mathbf{z}_{t_{i-1}}^2 + \Delta x^2}}{\sqrt{\Delta\mathbf{z}_{t_{i-1}}^2 + \Delta x^2}}, \tag{6.2}$$

where $\Delta\mathbf{z}_{t_i}$ denotes the spatial one-sided difference $\mathbf{z}(x+\Delta x) - \mathbf{z}(x)$ at time t_i.

Meanwhile, the loads on the insect wing, $p(x,t;k)$, result from two sources – the shed wake as well as external fluid disturbances, and are modeled by Theodorsen's equations [148] for lift forces, pressures and moments of a rigid airfoil characterized by oscillatory movement in the angle of attack $\alpha(k)$ and vertical position $h(k)$. Theodorsen's predictions agree quite well with experimental results [103, 28] compared to the predictions of quasi-steady thin airfoil theory, in which the lift coefficient is given by

$$C_L = \frac{2\pi}{U_\infty}\left[U_\infty\alpha + \dot{h} + \dot{\alpha}\left(\frac{1}{2} - \frac{b}{2}\right)\right]. \tag{6.3}$$

Theodorsen's model augments this quasi-steady lift coefficient (6.3) with the transfer function $C(k)$, derived by integrating forces generated by planar wake vorticity beyond the foil in inviscid, incompressible flow. Theodorsen's lift coefficient is given below, and we can distinguish added-mass and circulatory terms [103]

$$C_L = \frac{b\pi}{U_\infty^2}[U_\infty\dot{\alpha} + \ddot{h} - \frac{b^2}{2}\ddot{\alpha}] - \frac{2\pi}{U_\infty}C(k)\left[U_\infty\alpha + \dot{h} + \dot{\alpha}\left(\frac{1}{2} - \frac{b}{2}\right)\right]. \tag{6.4}$$

$C(k)$ is a quotient of Hankel functions of the second kind

$$C(k) = \frac{H_1^{(2)}(k)}{H_1^{(2)}(k) + iH_0^{(2)}(k)}, \tag{6.5}$$

Theodorsen's model will be used in our analysis for generating libraries of POD modes parametrized by reduced frequency k.

The first source of loads from the shed wake includes added masses and unsteady effects of the wing oscillation. These contributions are computed using the integral equation (6.6), which expresses the local chordwise loads $p(x,t;k)$ as a function of x and normal wing velocity $z_t(x,t)$. This expression is explicitly derived in [2] as an intermediate step to Theodorsen's result for the space-averaged global pressure coefficient.

$$\begin{aligned} p(x,t;k) = {} & \frac{2}{\pi}\mu U_\infty \sqrt{\frac{1-x}{1+x}} \int_{-1}^{1} \frac{z_t(\tilde{x},t)}{x-\tilde{x}} \sqrt{\frac{1+\tilde{x}}{1-\tilde{x}}} d\tilde{x} + \frac{1}{\pi}\mu b \int_{-1}^{1} z_{tt}(\tilde{x},t) L(x,\tilde{x}) d\tilde{x} \\ & + \frac{2}{\pi}\mu U_\infty [1 - C(k)] \sqrt{\frac{1-x}{1+x}} \int_{-1}^{1} z_t(\tilde{x},t) \sqrt{\frac{1+\tilde{x}}{1-\tilde{x}}} d\tilde{x}, \end{aligned} \tag{6.6}$$

where

$$L(x,\tilde{x}) = \log \frac{(x-\tilde{x})^2 + (\sqrt{1-x^2} - \sqrt{1-\tilde{x}^2})^2}{(x-\tilde{x})^2 + (\sqrt{1-x^2} + \sqrt{1-\tilde{x}^2})^2}.$$

The second source of chordwise loads are assumed to be external sinusoidal gusts characterized by different frequencies. We assume these gusts generate loads that resemble forces that would occur if the wing oscillates precisely at the characteristic frequencies of sinusoidal gust. Consequently, the loads resulting from external gusts are also computed using Theodorsen's model. This assumption approximates the aeroelastic effects of sinusoidal gust fields, since the forcing to the elastic beam includes these loads. This forced coupling between the elasticity model and the aerodynamic model addresses the absence of inertial effects in the Theodorsen framework. Specifically, we evaluate the following expressions for external loads, $p(x,t;k) = p_\alpha(x,t;k) + p_h(x,t;k)$,

$$\frac{p_\alpha(x,t;k)}{2b\mu\omega^2\alpha_0 e^{i\omega t}} = \left(\frac{U_\infty}{2} - \frac{i}{k}\right)\sqrt{1-x^2} + \frac{U_\infty}{2} x\sqrt{1-x^2} \tag{6.7}$$

$$- \frac{C(k)}{k^2}\sqrt{\frac{1-x}{1+x}} - \frac{1}{2k}(1 + 2x + 2C(k))\sqrt{\frac{1-x}{1+x}}, \tag{6.8}$$

$$\frac{p_h(x,t;k)}{2b\mu\omega^2 h_0 e^{i\omega t}} = \sqrt{1-x^2} - \frac{iC(k)}{k}\sqrt{\frac{1-x}{1+x}}, \tag{6.9}$$

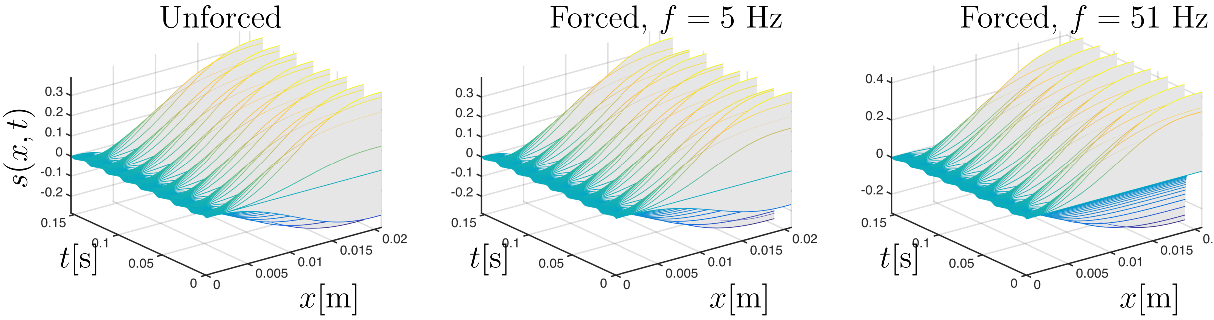


Figure 6.2: The strain dynamics for three different regimes, from left to right, (1) No gust disturbance, (2) low frequency gust disturbance $f = 5$ Hz, (3) high frequency gust disturbance $f = 51$ Hz. All three cases experience leading edge wing actuation at 26 Hz. Direct visualization of the strain dynamics offers no insight into the changing gust feedback between environments since inertial-elastic forces dominate over aerodynamic feedback forces.

derived by Postel and Leppert [127] in their analysis of Theodorsen's lift coefficient. The resulting loads from the gusts represent varying flow environments, hence, we generate gusts characterized by a wide range of reduced frequencies. Since the model wing and base flow remain unchanged, this is accomplished by varying f, hence k, and holding free-stream velocity and chord length fixed.

First, wing strain dynamics are visualized for three environments $i = 1, 2, 3$, all of which undergo the wing oscillation at 26 Hz. Environment 1 is unforced, and environments 2 and 3 undergo gust forcings of 5 Hz and 51 Hz, respectively. These frequencies correspond to reduced frequencies of $k \approx .03$ and $k \approx 0.3$, the first of which is quasi-steady, and the latter represents unsteady flow. Figure 6.2 displays the resulting strain dynamics which, by inspection, appear quite similar. The dominant spatial POD modes of strain, shown in Figure 6.3, confirm that spatial modes remain largely unaffected by the changing forcing term. Indeed, without additional inertial or added masses in the model, the various forced dynamics are still governed by the intrinsic spatial modes. This poses a difficulty for data-driven classifiers based on spatial POD modes. Indeed, preliminary runs of SRC on strain snapshots from the three environments yield classification accuracies of 30% on average, which is no better than random guesses. To remedy this, POD modes must incorporate the

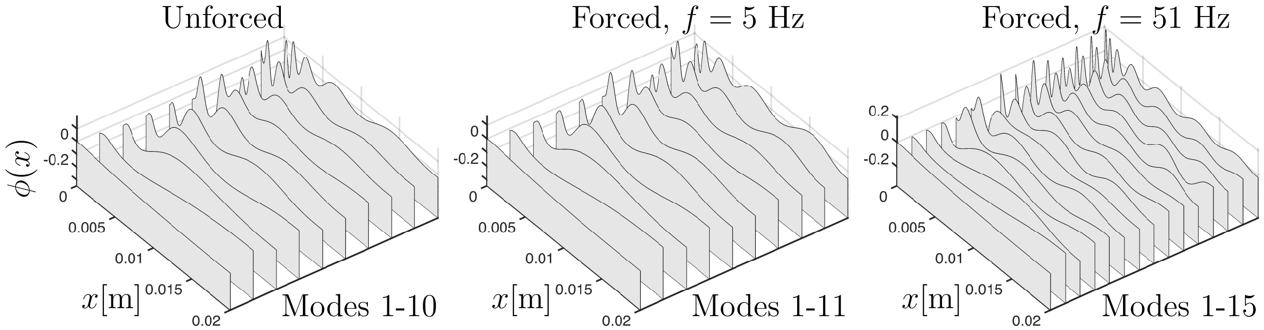


Figure 6.3: The similarity of dominant spatial POD modes of wing strain between regimes 1 to 3 (Figure 6.2) suggests that spatial POD modes are not ideal for the SRC classification task.

time dynamics of wing strain, as detailed in the following section.

6.3 Sparse frequency classification of wing strain

We introduce a sparse classification strategy that incorporates the time history of the strain signal through its frequency content. Indeed, insects and winged animals are thought to discriminate aerodynamic environments with only a few physical wing sensors that sample certain bandwidths of the feedback frequencies. Since gust forcing terms do not induce distinctive spatial modes, it is important to consider temporal strain dynamics, which are characterized by the jump discontinuities seen in Figure 6.2. These discontinuities are distributed across many temporal POD modes, posing difficulties in obtaining a low-rank truncated POD. Hence, we transform the time dynamics into the frequency domain using the discrete cosine transform (DCT), and adapt the POD library and sparse classifier to classify the transformed strain dynamics. Thus the input to the sparse classifier, $\hat{\mathbf{s}}$ (previously denoted $\mathbf{x}$), can be expanded in terms of frequency domain POD modes,

$$\mathbf{s}(x,\mathbf{t}) \xrightarrow{DCT} \hat{\mathbf{s}}(x,\mathbf{f}) = \sum_{j=1}^{r} a_j \hat{\boldsymbol{\phi}}_j(f) = \hat{\boldsymbol{\Psi}}\mathbf{a}. \tag{6.10}$$

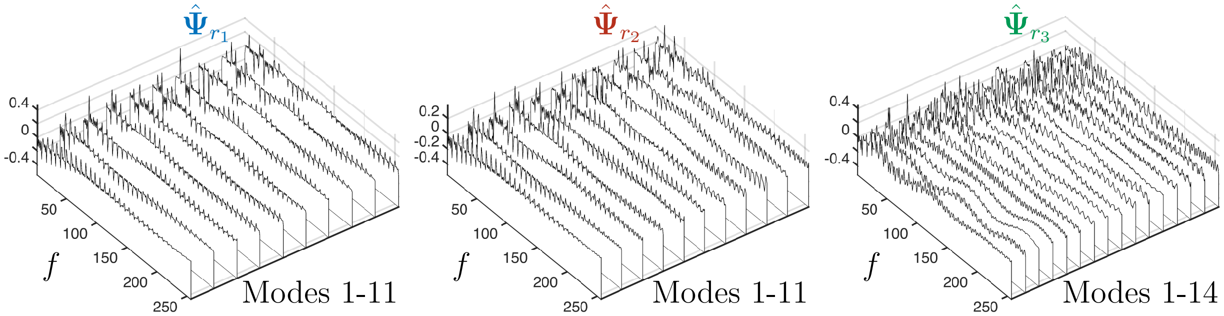


Figure 6.4: Dominant POD modes of frequency content for regimes 1 to 3 (left to right) concatenated into an overcomplete library. This representation is ideal for classification because the frequency content differs markedly across the three regimes.

The numerical simulation for environment i yields strain data $\mathbf{S}_i$ for m spatial gridpoints and n timesteps. Recall that the standard POD is performed on a data matrix whose columns are time snapshots, that is, its columns span the spatial direction. For frequency domain analysis, the strain data matrices $\mathbf{S}_i \in \mathbb{R}^{n\times m}$ are adjusted so that its columns span time dynamics

$$\mathbf{S}_i = [\mathbf{s}(x_1, \mathbf{t}) \mid \mathbf{s}(x_2, \mathbf{t}) \mid \ \ldots \ \mid \mathbf{s}(x_m, \mathbf{t})],$$

and the DCT applied to each column yields strain frequency content at every spatial location

$$\hat{\mathbf{S}}_i = [\hat{\mathbf{s}}(x_1, \mathbf{f}) \mid \hat{\mathbf{s}}(x_2, \mathbf{f}) \mid \ \ldots \ \mid \hat{\mathbf{s}}(x_m, \mathbf{f})].$$

Thus the procedure for building the POD library of frequency content remains unchanged

$$\hat{\mathbf{S}}_i = \hat{\mathbf{\Psi}}_{r_i} \mathbf{\Sigma}_{r_i} \mathbf{V}_{r_i}^T,$$

where hatted notation indicates that POD modes span frequency content. Figure 6.4 illustrates the dominant POD modes of the strain data's temporal frequency content for the three regimes. The frequency content reveals substantial differences between the aerodynamic regimes that facilitate robust, accurate classification.

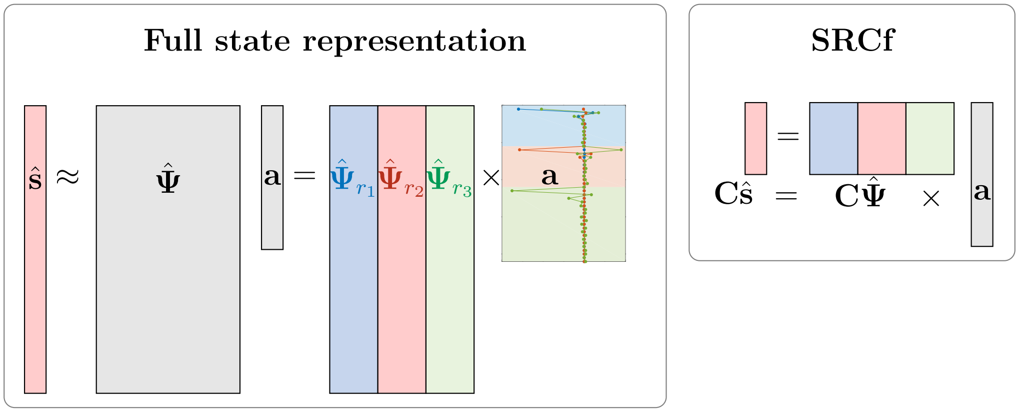


Figure 6.5: SRCf promotes sparsity of the coefficients $\mathbf{a}$ to help identify the correct originating environment i using sparse frequency measurements $\mathbf{C}\hat{\mathbf{s}}$. The left plot contains 3 actual solution vectors. The blue line is the solution $\mathbf{a}$ for a test vector from regime (1), red from (2), and green from (3). Each line correctly identifies its originating regime with its largest magnitude nonzero components.

6.3.1 Sparse approximation for classification (SRC)

Given the frequency adjusted POD library, the sparse approximation procedures for classification are analogous to those in the previous chapter. As before, data-driven modes $\hat{\boldsymbol{\Psi}}_{r_i}$ are stacked into the frequency POD library $\hat{\boldsymbol{\Psi}}$, and a new state is classified from the POD library coefficients. The input to the classifier is the frequency content at one spatial location on the wing, $\hat{\mathbf{s}}$. The ℓ_1 constrained approximation for the full state is then given by

$$\mathbf{a} = \arg\min_{\tilde{\mathbf{a}}} \|\tilde{\mathbf{a}}\|_1 \text{ subject to } \|\hat{\boldsymbol{\Psi}}\tilde{\mathbf{a}} - \hat{\mathbf{s}}\|_2 < \epsilon. \tag{6.11}$$

Note that the error tolerance ϵ must be chosen carefully to balance the trade-off between approximation and classification.

Suppose only a few frequencies (fewer than the number of columns in the library) are discretely sampled and stored within a vector $\mathbf{y} = \mathbf{C}\hat{\mathbf{s}}$, which satisfies

$$\mathbf{y} = \mathbf{C}\hat{\boldsymbol{\Psi}}\mathbf{a}.$$

This is an underdetermined set of equations with multiple solutions, and the sparsest one is given by the following optimization

$$\mathbf{a} = \arg\min_{\tilde{\mathbf{a}}} \|\tilde{\mathbf{a}}\|_1 \text{ subject to } (\mathbf{C}\hat{\mathbf{\Psi}})\tilde{\mathbf{a}} = \mathbf{C}\hat{\mathbf{s}}.$$

The SRC schemes are illustrated in Figure 6.5. As before, the strength of $\mathbf{a}$'s components will determine the subset of regimes to which the strain frequency signal belongs. Because the classifier now operates in the frequency domain, SRC in the frequency domain is abbreviated as *SRCf*.

6.3.2 Optimal measurement selection for SRCf

Engineering and biological applications typically benefit from optimized measurement strategies for decision-making and estimation tasks. Recall that ℓ_1 norm constrained SRC permits sparse input sampling using sensor placement methods which identify measurement locations optimized for regression within the POD basis. Since *locations* correspond to the frequency bandwidths accessible for measurement, we refer to this optimized measurement selection as frequency selection.

Measurement Selection

As we have seen in previous chapters, measurements can be optimized for least-squares reconstruction in a single low-rank POD basis. To generalize these methods for ℓ_1 classification in a library of modes, we first obtain a single low-rank POD basis for all categories of data. For categorical data, a joint POD of all categories yields fewer modes than a concatenated library of POD modes from each consecutive category. Importantly, both are assumed to characterize the same active subspace or column space, since a joint POD compactly represents redundant features shared between categories in a POD library. Explicitly, we construct a

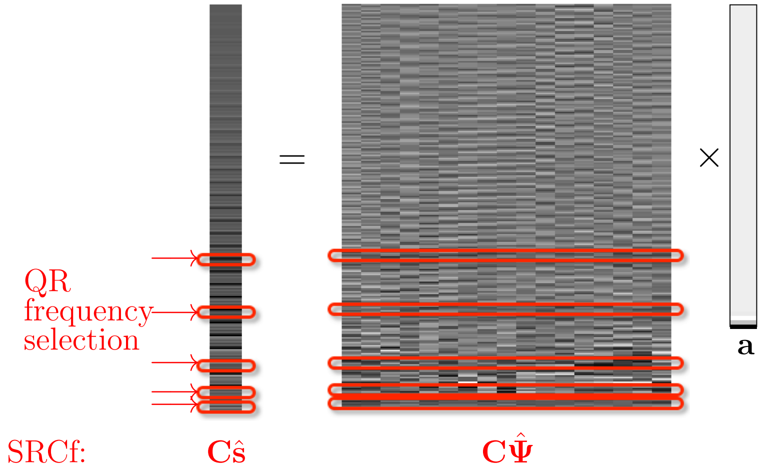


Figure 6.6: This figure illustrates the action of the selection operator $\mathbf{C}$ to the sparse approximation problem and converts the overdetermined SRCf into the underdetermined compressed SRCf problem.

joint POD using strain data from all flow categories,

$$[\hat{\mathbf{S}}_1 \mid \hat{\mathbf{S}}_2 \mid \ \ldots \hat{\mathbf{S}}_c] = \mathbf{\Theta\Sigma V}^T, \tag{6.12}$$

and truncate the left singular vectors to retain only the first r dominant POD modes. This is advantageous for engineering the fewest measurements possible for sparse classification, since

$$r < \sum_{i=1}^{c} r_i.$$

Subsequently, sparse library coefficients $\mathbf{a}$ may be recovered from selected measurements of the full signal, where we denote the measurement selection operator by $\mathbf{C}$

$$(\mathbf{C}\hat{\mathbf{\Psi}})\mathbf{a} \approx \mathbf{C}\hat{\mathbf{s}}. \tag{6.13}$$

For the moment, we proceed as though the unknown state $\hat{\mathbf{s}}$ is reconstructed (not classified) in $\mathbf{\Theta}_r$, the POD feature space spanning all classes. Since we cannot access the full $\hat{\mathbf{s}}$, it cannot be recovered from its POD coefficients $\boldsymbol{\Theta}_r^T\hat{\mathbf{s}}$.

Instead, the goal is to optimize $\mathbf{C}$ for approximating $\hat{\mathbf{s}}$ from $\mathbf{y} = \mathbf{C}\hat{\mathbf{s}}$ as follows

$$\mathbf{C}_\star = \underset{\mathbf{C}}{\arg\min} \, \|\hat{\mathbf{s}} - \boldsymbol{\Theta}_r(\mathbf{C}\boldsymbol{\Theta}_r)^{-1}\mathbf{y}\|_2.$$

As before, we compute r optimal point measurement indices from the QR factorization pivots of the dominant dynamical modes $\boldsymbol{\Theta}_r$

$$\boldsymbol{\Theta}_r^T\mathbf{C}^T = \mathbf{QR}, \tag{6.14}$$

which is described in Algorithm 2. More generally, QR measurement selection can easily be adapted to incorporate measurement location constraints which are commonly encountered in engineering applications. Undesirable measurement locations from a practical standpoint can simply be omitted from the QR pivoting algorithm by omitting the appropriate rows (measurements) from the input $\boldsymbol{\Theta}_r$. More complex preferences can be implemented by multiplying the input modes by application-specific weight matrices; this is the focus of ongoing work.

6.4 Results and discussion

In this section sparse approximation simulations are designed to detect flow-induced strain patterns across various flow environments, frequency samples and spatial correlations. Importantly, the advantages of sparse over conventional least squares approximation and trained QR row selections over random samples in SRCf are demonstrated. These patterns are of particular importance in engineering applications that require collection of strain data optimized for environment detection. These tests also stress the importance of sensor noise in these applications by demonstrating robust sparse approximation performance over a range of increasing noise levels in the strain dynamics.

Here machine learning terminology is used to describe the data. Single location frequency

Method	**Figure**	**Input dimension**	**Sensors**	**Gust frequencies**	**No. library modes**
ℓ_1 approximation	8, §5.1	256	–	Ø,5,51	36
Least squares	9, §5.2	256	–	Ø,5,51	36
SRCf	10, §5.3	5	QR	Ø,25,50,75,100	75
	11, §5.3	10	QR	Ø,25,50,75,100	75
	12, §5.4	35	Random	Ø,25,50,75,100	75

Table 6.2: Summary of numerical experiments and sparse classifiers, where Ø refers to the flow environment with no gust forcing. All flow environments exhibit the dominant 26 Hz frequency of wing oscillation. Furthermore, the number of gust frequencies equals the total number of classes considered.

content are called observations or signals, to disambiguate from system states that may refer to snapshots in time. Observations to be classified, $\hat{\mathbf{s}}$ or $\mathbf{y} = \mathbf{C}\hat{\mathbf{s}}$, are called inputs to the classifier. As a pre-processing step, noise is added to each column of the untransformed input observations in the following order.

1. Normalize each column so that $\|\mathbf{s}(x_j, \mathbf{t})\|_2 = 1$ for all $j = 1 \to m$ gridpoints,
2. Add zero-mean Gaussian noise so $\mathbf{S} = \mathbf{S} + \boldsymbol{\zeta}$, $\boldsymbol{\zeta} \sim N(0, \eta^2)$, and
3. Transform $\mathbf{S}$ into frequency domain $\hat{\mathbf{S}}$: $\hat{\mathbf{s}}(x_j, \mathbf{f}) = DCT[\mathbf{s}(x, \mathbf{t})]$.

Thus sensor noise is added to the untransformed normalized signal in the space-time dimension as would be the case in engineering applications, and classifier inputs are normalized to facilitate comparing the effect of noise across different environmental regimes. Note that we do not normalize or add noise to the transformed signal $\hat{\mathbf{s}}$ in the frequency domain since wing strain measurements are noisy when collected (sensor noise), not when processed via the frequency domain (process noise). Table 6.2 provides a summary of the numerical experiments performed and the figures produced for this section.

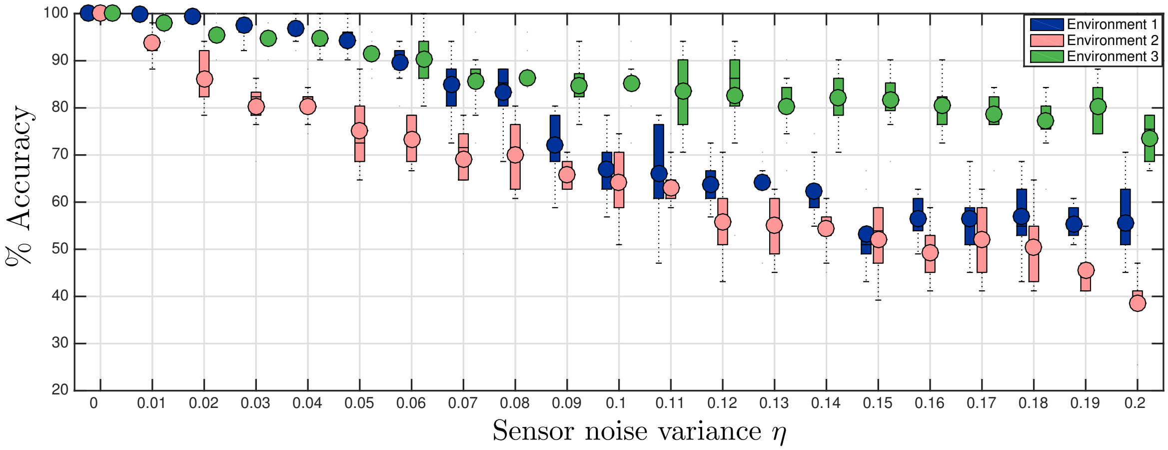


Figure 6.7: The cross-validated performance of ℓ_1 approximation across environments 1-3 is shown here for sensor noise of increasing variance η. Red, blue and green represent input observations from environments 1,2 and 3, respectively. High-frequency gusts from regime 3 are more robustly distinguished from the other two regimes.

6.4.1 Feedback identification from full state

We demonstrate sparse flow feedback identification between the three learned flow environments in section 6.2, and test it on randomly selected subsets of strain dynamics from each flow environment. The environments are chosen to highlight the extremes of gust frequencies - low feedback of $f = 5$ Hz in environment 2, high feedback of $f = 51$ Hz in environment 3, and no feedback in regime 1. All three share the 26 Hz frequency signature expressed by wing oscillation. It is expected that some environments will be identified with higher accuracy than the others, and strain dynamics from each environment are sparsely classified separately to reveal this structure in the data.

For each flow environment i, the tests are conducted using tenfold cross validation. Ten percent of vectors in $\hat{\mathbf{S}}_i$ are randomly selected to be test inputs of strain dynamics, and the remaining 90% of the columns are used to train the POD library and optimal sensors, meaning the POD modes from each environment are trained from 500 observations. The POD library $\hat{\boldsymbol{\Psi}}$ consists of the stacked dominant POD modes of environments 1,2 and 3 shown in Figure 6.4. After training the POD library, test frequency dynamics $\hat{\mathbf{s}} \in \mathbb{R}^{256}$

are classified using ℓ_1 constrained approximation of the overdetermined linear system(6.11). The performance is investigated across increasing levels of sensor noise, and the classification accuracies across all the tests for one level of sensor noise are displayed as one box and whisker in the subsequent plots. The mean, 25th quartile, and 75th quartile of the classification performance distribution are displayed as circles, the bottom, and top edges of the boxes, respectively. Whiskers extend to ±2.7 standard deviations of the result distribution, and any data outside this range is displayed as a small outlier point.

Box and whisker plots of classification accuracy for test dynamics from environments 1-3 are shown in Figure 6.7. ℓ_1 constrained approximation achieves perfect 100% classification accuracy in the noiseless case for all three environments, and even at levels of 20% sensor noise, the performance remains impressively higher than chance or 33.3%, the scenario in which the classifier chooses between the three environments uniformly at random. Interestingly, as sensor noise increases, the identification of high-frequencies in environment 3 is more successful than the identification of environments 1 and 2. Environment 3 in particular contains high frequency content and higher amplitude spatial modes (Figure 6.2) that are amplified in the frequency domain analysis.

These results are quite sensitive to the choice of ℓ_2 error tolerance ϵ. In these computations, ϵ is set to some multiple of the least squares solution error. That is, given the cheaply computed least squares solution $\mathbf{a}_{ls} = \arg\min_{\tilde{\mathbf{a}}} \|\hat{\mathbf{\Psi}}\tilde{\mathbf{a}} - \hat{\mathbf{s}}\|_2$, $\epsilon_{ls} = \min_{\tilde{\mathbf{a}}} \|\hat{\mathbf{\Psi}}\tilde{\mathbf{a}} - \hat{\mathbf{s}}\|_2$. In practice, setting $\epsilon = \epsilon_{ls}$ in the noiseless case and relaxing the tolerance to $\epsilon = 1.2\epsilon_{ls}$ in the noisy case works quite well. The tolerance appears to depend on the noise level η and ϵ_{ls}, and there may exist an optimal choice based on the two parameters requiring further investigation.

6.4.2 Sparse (ℓ_1) vs. least squares (ℓ_2) approximation

It is instructive to overlay in Figure 6.8 results from classification using the least squares solution. The comparison demonstrates the clear advantage of using ℓ_1 norm minimization of library coefficients over ℓ_2 minimization of the residual (least squares), particularly in the presence of sensor noise, although least-squares achieves ℓ_1 accuracy in the noiseless

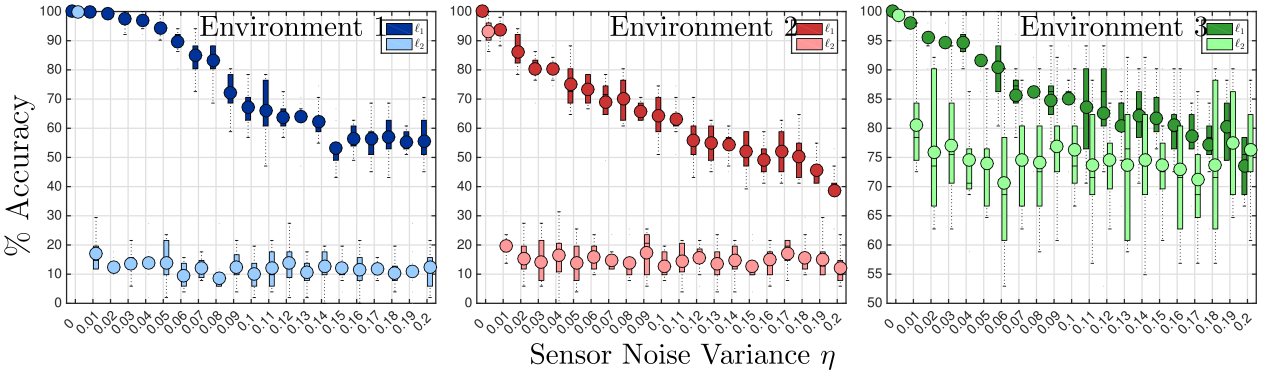


Figure 6.8: The comparison of ℓ_1 approximation to least squares approximation of the library coefficients demonstrates the power of the sparsity promoting ℓ_1 norm constraint for subset selection.

case. The identification of environment 3 is once again an exception due to high-frequency fitting that is not present in the other two, however, ℓ_1 constrained approximation always outperforms least squares. Therefore the sparsity promotion of solution library coefficients is essential for effective classification in the discriminating frequency domain.

6.4.3 SRCf with QR selection

The resolution of strain measurements in time is directly informed by the optimized sampling of frequency content for classification. The strain dynamics from different flow environments by construction are linearly separable by flow feedback frequency. The precise discriminating frequencies are unknown in the training stage but are easily determined from QR selection with the joint POD of all strain dynamics. The following simulation compares the performance of SRCf with QR row selection over random samples in the frequency domain, confirming the advantage of QR selection.

QR selection is obtained from a richer training dataset of strain dynamics consisting of forced disturbances at 25 Hz, 50 Hz, 75 Hz and 100 Hz, in addition to the no forcing scenario. There are now five environments considered by the classifier, a more difficult task than the former. As before, the tested dynamics do not consist of the same random 90% of the data

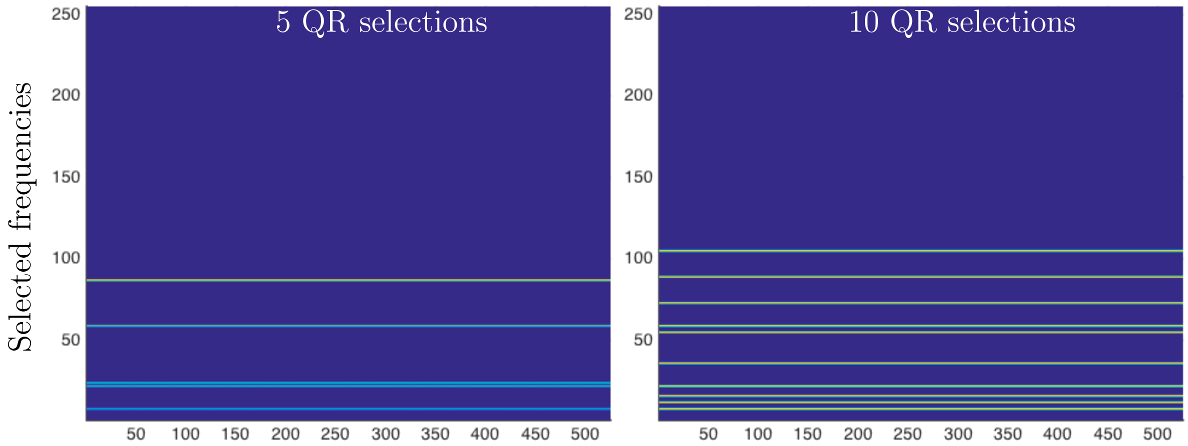


Figure 6.9: Ensemble of all QR frequency selections from nearly five-hundred instances of cross-validation, that is, QR selection using POD modes trained from five-hundred random observations. Note that optimal sampling locations are stable and experience no variation across different training sets.

used for training POD library modes and QR selection. The five QR selected measurements for the five environment case in Figure 6.9 appear to cluster around the discriminating disturbance frequencies and even appear to discriminate the flow forcing at 25 Hz from wing oscillation at 26 Hz. The measurements do not exactly lie at these numbers because the DCT does not span whole number frequencies.

The performance of SRCf shown in Figure 6.10 demonstrates the advantage of QR selection of strain dynamics over random sampling. Up to a 35% increase in classification accuracy is achieved with only five samples (1.95% of the 256 element vector of strain dynamics). In addition, QR selection is more robust to sensor noise than an even larger number of random samples, although the accuracy gain drops off near sensor noise of level $\eta = 0.2$ that comprises 20% of the untransformed signal's magnitude. The strain dynamics are saturated by noise at this higher noise level where SRCf performance drops off to randomized classification accuracy.

Interestingly, doubling the number of QR selections does not improve the accuracy, although doubling the number of random samples improves random performance as expected.

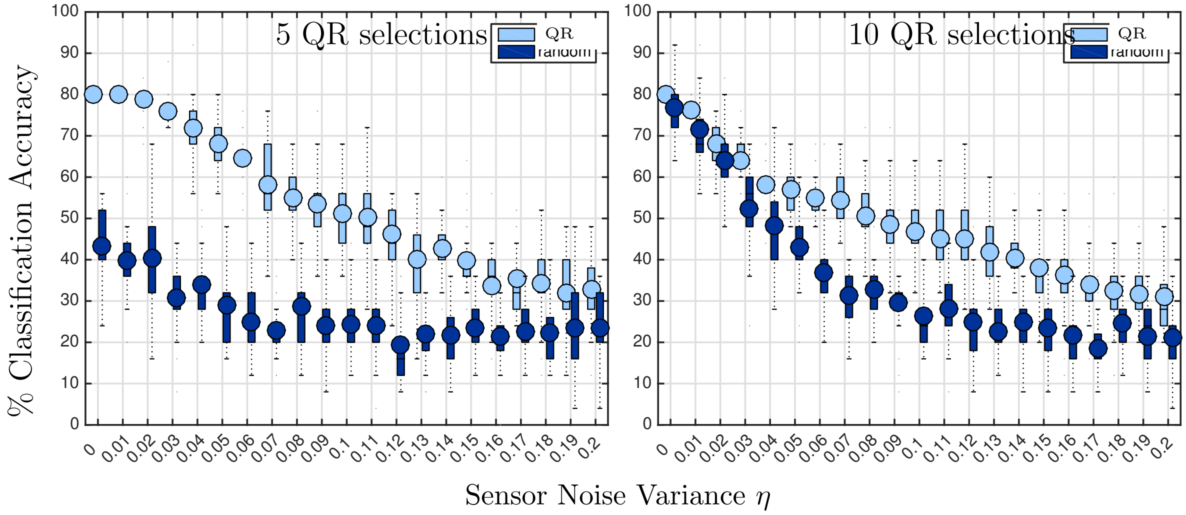


Figure 6.10: **SRCf Accuracy**: This comparison of QR selected frequencies with random frequency samples demonstrates that even a 2% (5 row) QR selection outperforms a 14% random row selection (35 rows) for classifying five environments for increasing sensor noise.

The results indicate that cross-environment variation is well characterized by the QR selected measurements which are optimized for library approximation with $\hat{\mathbf{\Psi}}$. Furthermore, increasing the number of random samples to 35 (Figure 6.11) does not achieve the same robust accuracy drop-off seen with QR selections.

6.4.4 SRCf with spatial bias

A related inquiry concerns the optimal spatial locations along the wing to capture strain dynamics. These locations would correspond to the physical placement of each strain sensor and campaniform sensilla on insect wings. Our previous simulations ignored any spatial correlation by testing classification for frequency content from random locations along the chord (columns of $\hat{\mathbf{S}}$). We now introduce a spatial bias that draws strain frequency dynamics at random from Gaussian distributions centered at different regions of the wing chord and training with the remainder, using the same data considered previously in section 6.4.3. This essentially tests columns $\hat{\mathbf{s}}$ that are adjacent in the data matrix $\hat{\mathbf{S}}$. The different regions of the chord considered - the leading edge, 1/4-chord, midchord, 3/4-chord, and trailing - are

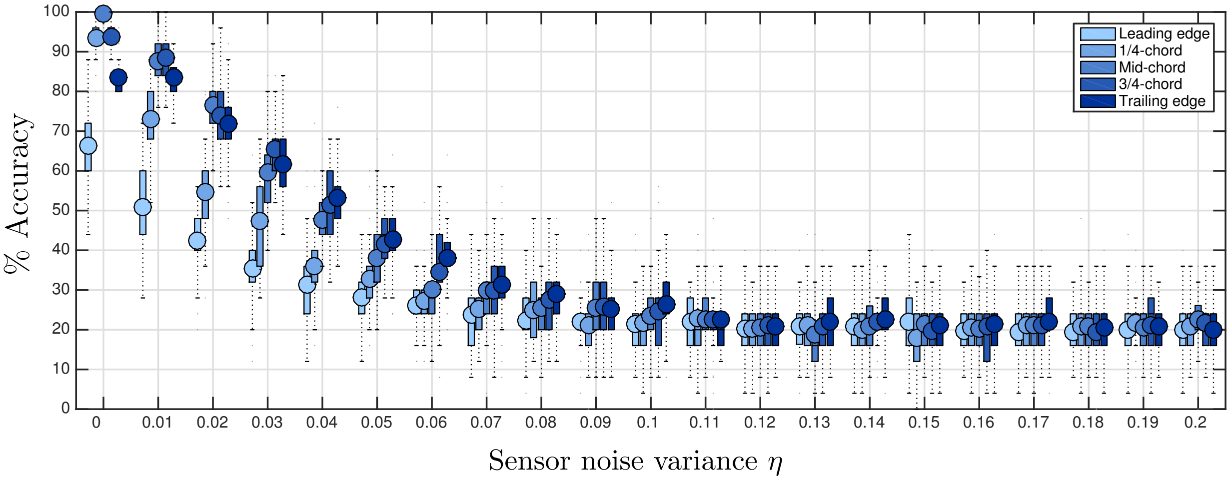


Figure 6.11: This plot compares classification accuracy for sensors distributed around (left to right) leading edge, 1/4 chord, mid-chord, 3/4-chord and trailing edge for identically increasing sensor noise. Each error bar represents cross-validated classification accuracy across observations randomly chosen from a Gaussian distribution centered around a certain area of the chord, sampled from a randomly chosen environmental regime out of the five total regimes. Classification accuracy increases across the chord with accuracy increasing near the trailing edge.

important in the study of both insect wings and airfoils. The SRCf classifier is used with 35 random frequency samples in the frequency domain.

The expected performance profile for increasing sensor noise is observed in the results of Figure 6.11, however there is an obvious classification advantage towards the trailing edge of the wing. This trend is consistent with downstream amplification of flow-induced strain towards the trailing edge, where the wing is most prone to deflection from the free boundary (6.1) and flexural stiffness is lowest. However, this finding deviates from observed campaniform sensilla locations that tend to be distributed away from the trailing edge. There are many possible explanations - the trailing edge is dominated by membrane surface without venal or neural conduits for the sensilla, or the non-stiff trailing edge of the wing may be prone to amplified flow process noise. Alternatively, the ensemble of campaniform sensilla may be sufficient to aggregate strain dynamics to amplify flow features without resorting to trailing edge sensors. Both scenarios are well worth examining in future study.

6.5 Conclusions

This work develops a data-driven framework for flow environment identification based on supervised learning from simulated wing strain dynamics in the frequency domain. This is done using sparse classification in an overcomplete library of POD modes, which is assembled by simulating strain dynamics from each expected flow environment. Then, the problem of identifying the flow environment reduces to a classification task between the sets of POD modes in the library. The subset selection is accomplished by ℓ_1 regularized, sparse approximation of POD library coefficients from input frequency dynamics. This ℓ_1 regularization is extremely accurate and mitigates contributions from noise when compared to least squares approximation – even with sensor noise at 20% of the input signal's magnitude, sparse classification is slightly biased towards the correct originating environment. Furthermore, SRCf with strategically subsampled signals is shown to be effective, with discriminating frequency measurements selected by the pivoted QR factorization of POD modes. In this manner, sparse classification of strain frequencies facilitates environment identification from single wing locations, providing insight into possible sparse encoding strategies employed by strain sensing neurons on insect wings. Moreover, our approach may be applied to sensor-equipped feedback systems in flight control, which is an active area of research.

6.5.1 Control Implications

Downstream environment identification from sensor data is an important problem in closed-loop control where sensors are expensive and measurements are corrupted by noise. Spatial strain mode shapes do not vary significantly across environments and require many more physical strain sensors to complete spatial knowledge. Frequency domain analysis of wing strain can help move toward autonomous flight technologies. Furthermore, the design of frequency samples chosen specifically to discriminate certain environments or frequencies greatly reduces the dimension of decision space and helps mitigate uncertainty in sensor measurements. Strain sensors and gauges are already being used to characterize stresses and guide wing and fin design in experiment [87]. As an added functionality they may assist in flow characterization using similar data-driven dimension reduction concepts as presented here. Our frequency distinction framework is particularly attractive for feedback analysis for controller decisions. Given suitable libraries of candidate frequency dynamics, the sparse classification framework can be efficiently implemented in flight controllers, especially when the decision space consists of only a few strategically selected frequencies. For example, controllers may mitigate undesirable frequencies encountered in flight such as the erratic high frequency content of turbulence and low frequencies induced by strong winds. Although we have not built a controller in this work, we note that convex optimization strategies, such as the one underlying sparse classification, are commonly employed in control applications.

6.5.2 Biological Implications

There are a number of intriguing biological conjectures that the current results help address. Specifically, the machine learning architecture of learned dictionaries and sparse sampling both seem to have advantageous properties in the context of flight. The learned libraries encoded from our numerical simulations suggest that insects possess a similar, innate library for flight dynamics. Indeed, insects at birth simply begin flying without a lengthy training stage, suggesting that flight dynamics, or dictionaries, are stereotyped behavior in the mechanosensory flight system [48, 68, 57]. Thus at birth, a low-dimensional representation of mechanosensory codes are genetically inherited. The sparse and stereotyped placement of campaniform sensilla on the wings of a hawkmoth also suggest that sparse sampling does indeed occur for helping guide flight dynamics and control protocols. Of course, we have made no connection to control, but we certainly have demonstrated the tremendous bio-inspired advantages of sparsity and learned libraries.

Chapter 7

SENSOR AND ACTUATOR PLACEMENT FOR OPTIMAL CONTROL

In this chapter we consider a generalization of linear dynamical systems with outputs

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} & \mathbf{x} &\in \mathbb{R}^n \\ \mathbf{y} &= \mathbf{C}\mathbf{x}, & \mathbf{y} &\in \mathbb{R}^p, \end{aligned}$$

to linear dynamical systems with both inputs $\mathbf{u}$ and outputs $\mathbf{y}$

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} & \mathbf{x} &\in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^q \\ \mathbf{y} &= \mathbf{C}\mathbf{x}, & \mathbf{y} &\in \mathbb{R}^p, \end{aligned} \tag{7.1}$$

This is a natural extension of the sensor placement problem considered in earlier chapters to combined sensor and actuator placement. Specifically, we are still selecting point observations of the state $\mathbf{x}$ using measurement operator $\mathbf{C}$, and also selecting which states to actuate using $\mathbf{B}$. Point configurations for a q-in p-out system require that sensing and actuation matrices $\mathbf{C} \in \mathbb{R}^{p\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times q}$ are structured in the following way

$$\begin{aligned} \mathbf{C} &\triangleq \begin{bmatrix} \mathbf{e}_{\gamma_1} & \mathbf{e}_{\gamma_2} & \dots & \mathbf{e}_{\gamma_p} \end{bmatrix}^T \\ \mathbf{B} &\triangleq \begin{bmatrix} \mathbf{e}_{\beta_1} & \mathbf{e}_{\beta_2} & \dots & \mathbf{e}_{\beta_q} \end{bmatrix}, \end{aligned} \tag{7.2}$$

where $\mathbf{e}_i$ is the n-dimensional vector of zeros with a unit entry at the ith component. As before, $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_p\} \subset \{1, \dots, n\}$ denotes the index set of sensor locations with cardinality $|\boldsymbol{\gamma}| = p$. Similarly, actuator selection indices are given by $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_q\}$.

7.1 Motivation

The feedback control of any system relies on the quality of actuation input and measurement output, and in many cases, sensing and actuation capabilities are limited and expensive. This is especially true in complex systems such as wings and motors interacting with fluid flows and turbulence. When discretized, the added challenge of searching through the grid for appropriate placements is computationally intractable. A sensible strategy is to place sensors along the most observable directions in state space, and actuators along the most controllable directions. Indeed, there is a vast body of literature on placements based on observability [4, 79, 44] and controllability [142, 14]. Existing computational methods can determine continuous (off-grid) placements for using gradient descent [43] or variable-transformed convex optimization [44], as well as grid-based placements using greedy methods [131] and convex optimization [18, 84]. Specific control objectives include minimizing the H_2 norm or H_∞ norms of resulting configurations [43, 119]. There are placement methods conceptually similar to our proposed approach. For instance, one can exploit balanced model reduction to more efficiently evaluate these objectives [119], or investigate connections between POD-based sensor placement and system observability [4]

Clearly, any placement of either sensors or actuators must take into consideration the input output dynamics of the system, and in general, these are not well-characterized by global POD mode analysis. To generalize the pivoting approach formulated in previous sections, we take into account the most observable and controllable directions in a system. Our approach relies on balanced model reduction [113, 134, 166], in which modes are hierarchically ordered by degree of observability and controllability. Here, the QR pivoting algorithm samples the low-rank balancing transformations of the system to find maximally observable and controllable locations in state space. The resulting locations correspond to near-optimal point sensor and actuator configurations. The quality of our optimized configurations is evaluated using the H_2 norm and Hankel singular values of the resulting system, both of which measure the input-output energy in a system. We demonstrate our method on random state space systems and the linearized Ginzburg-Landau equation with stochastic disturbances. The resulting sensor and actuator configurations reproduce known optimal configurations at a fraction of the computational cost associated with expensive gradient optimization methods.

7.2 Background on balanced model reduction

This section details the dimensionality reduction method that generalizes our QR pivoting approach to control. The primary tool at our disposal is balanced model reduction, which seeks a rank r biorthogonal basis that characterizes the most observable and controllable directions in state space. First, we define gramians that quantify the degrees of observability and controllability. This motivates a *balanced* low-rank representation of these gramians, in which the most observable and controllable states are hierarchically ordered by their contribution to the original system. Next, we describe this so-called *balanced truncation* and its computational implementation, the results of which will then be used for QR-optimized sensor and actuator placement.

7.2.1 Observability and controllability

The degrees of observability and controllability for the state-space system (7.1) are quantified by the observability gramian

$$\mathbf{W}_o = \int_0^\infty e^{\mathbf{A}^* t}\mathbf{C}^*\mathbf{C}e^{\mathbf{A}t}dt \tag{7.3}$$

and controllability gramian,

$$\mathbf{W}_c = \int_0^\infty e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^*e^{\mathbf{A}^* t}dt, \tag{7.4}$$

respectively. The gramians $\mathbf{W}_o$ and $\mathbf{W}_c$ define ellipsoids in state-space that contain all the states observable from a given output, and all the states reachable from a given input. In other words, the maximal energy output from a given initial condition $\mathbf{x}_0$ is quantified by the observability gramian

$$\max \|\mathbf{y}\|^2 = \mathbf{x}_0^* \mathbf{W}_o \mathbf{x}_0. \tag{7.5}$$

Likewise, the minimal energy required to steer a system to a given state is defined by the inverse controllability gramian

$$\min \|\mathbf{u}\|_2 = \mathbf{x}_0^* \mathbf{W}_c^{-1} \mathbf{x}_0. \tag{7.6}$$

Note that the gramians define ellipsoids as a function of sensing and actuation ($\mathbf{C}$ and $\mathbf{B}$). Intuitively, this suggests designing sensors and actuators to optimize some scalar measure of these gramians. However, the gramians are high-dimensional and in general, are not equal, since the observable and controllable subspaces of the system may be different. Moreover, dimensionality reduction with POD is insufficient because it does not accommodate these different subspaces.

The primary goal of balanced model reduction is to find a coordinate transformation $\mathbf{x} \approx \boldsymbol{\Psi}_r \mathbf{a}$ giving rise to a related system $(\mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r)$ with similar input–output characteristics, in terms of a state $\mathbf{a} \in \mathbb{R}^r$ with many fewer degrees of freedom, $r \ll n$. Note that $\mathbf{u}$ and $\mathbf{y}$ are the same as in Eq. (7.1), even though the system state has been reduced. Instead of ordering modes based on energy, as in the POD, it is possible to determine a hierarchy of modes that are most controllable and observable, therefore capturing the most input–output information [113]. These modes give rise to *balanced* models, equally weighting the controllability and observability of a state via a coordinate transformation that makes the controllability and observability Gramians equal and diagonal. These models have been quite successful, although computing balanced models may be computationally expensive.

7.2.2 Balanced truncation and the balancing transformation

Balanced model reduction is based on finding a coordinate transformation $\mathbf{x} = \boldsymbol{\Psi}\mathbf{a}$ that hierarchically orders the states in $\tilde{\mathbf{x}}$ in terms of their ability to capture the input–output characteristics of the system, as quantified by the controllability and observability Gramians. In the seminal paper by Moore in 1981 [113], he showed that it was in fact possible to find a coordinate system where the controllability and observability Gramians are equal and diagonal. This results in the balanced model:

$$\begin{aligned} \dot{\mathbf{a}} &= \boldsymbol{\Phi}^*\mathbf{A}\boldsymbol{\Psi}\mathbf{a} + \boldsymbol{\Phi}^*\mathbf{B}\mathbf{u} \qquad && \mathbf{z} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^p \\ \mathbf{y} &= \mathbf{C}\boldsymbol{\Psi}\mathbf{a}, && \mathbf{y} \in \mathbb{R}^q \end{aligned} \tag{7.7}$$

where the balancing modes $\boldsymbol{\Psi}$ are called *direct* modes and $\boldsymbol{\Phi}^* = \boldsymbol{\Psi}^{-1}$ are called the *adjoint* modes.

The balanced state $\mathbf{z}$ is then *truncated*, keeping only the first $r \ll n$ most jointly controllable and observable states in $\mathbf{a}$, so that $\mathbf{x} \approx \boldsymbol{\Psi}_r\mathbf{a}$, resulting in the following *balanced truncation* model [113]:

$$\begin{aligned} \dot{\mathbf{a}}_r &= \boldsymbol{\Phi}_r^*\mathbf{A}\boldsymbol{\Psi}_r\mathbf{a}_r + \boldsymbol{\Phi}_r^*\mathbf{B}\mathbf{u} \qquad && \mathbf{a}_r \in \mathbb{R}^r, r \ll n \\ \mathbf{y} &= \mathbf{C}\boldsymbol{\Psi}_r\mathbf{a}_r. \end{aligned} \tag{7.8}$$

Note that this reduced-order model is formulated in terms of the original system matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and the first r columns of the direct and adjoint modes, $\boldsymbol{\Psi}_r$ and $\boldsymbol{\Phi}_r$, respectively.

The controllability and observability Gramians each establish an inner product on state space in terms of how controllable or observable a given state is, respectively. As such, Gramians depend on the particular choice of coordinate system and will transform under a change of coordinates. In the coordinate system $\mathbf{a}$, the controllability Gramian becomes:

$$\tilde{\mathbf{W}}_c = \boldsymbol{\Phi}^*\mathbf{W}_c\boldsymbol{\Phi}. \tag{7.9}$$

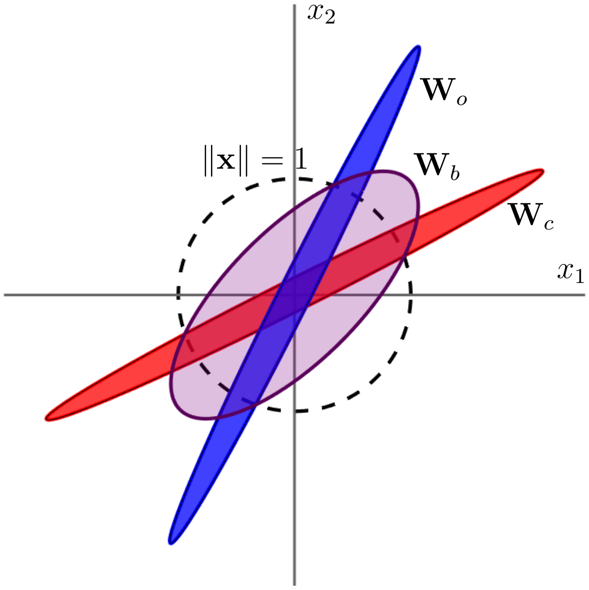


Figure 7.1: Illustration of balancing transformation on Gramians. The reachable set with unit control input is shown in red, given by $\mathbf{W}_c^{1/2}\mathbf{x}$ for $\|\mathbf{x}\| = 1$. The corresponding observable set is shown in blue. Under the balancing transformation $\mathbf{\Psi}$, the Gramians are equal, shown in purple. Reproduced with permission from Brunton and Kutz [30].

The observability Gramian transforms similarly:

$$\tilde{\mathbf{W}}_o = \mathbf{\Psi}^*\mathbf{W}_o\mathbf{\Psi}. \tag{7.10}$$

The coordinate transformation $\mathbf{\Psi}$ that makes the controllability and observability Gramians equal and diagonal,

$$\tilde{\mathbf{W}}_c = \tilde{\mathbf{W}}_o = \mathbf{S}, \tag{7.11}$$

is given by the matrix of eigenvectors of the product of the Gramians $\mathbf{W}_c\mathbf{W}_o$ in the original coordinates:

$$\tilde{\mathbf{W}}_c\tilde{\mathbf{W}}_o = \mathbf{\Phi}^*\mathbf{W}_c\mathbf{W}_o\mathbf{\Psi} = \mathbf{S}^2 \quad \Longrightarrow \quad \mathbf{W}_c\mathbf{W}_o\mathbf{\Psi} = \mathbf{\Psi}\mathbf{S}^2. \tag{7.12}$$

The diagonal entries of $\mathbf{S}$ are the *Hankel singular values* of the system, where the ith singular value represents the input-output energy contribution of the ith balancing mode pair. Subsequently, modes with small contribution are truncated to construct a balanced truncation that approximates the full system dynamics.

Computational algorithms

Computing this eigendecomposition relies on computing or approximating both gramians. The gramians are defined by the infinite time-integration for stable systems, but are also the unique solutions of the following Lyapunov equations

$$\mathbf{A}\mathbf{W}_c + \mathbf{W}_c\mathbf{A}^* + \mathbf{B}\mathbf{B}^* = 0 \tag{7.13a}$$

$$\mathbf{A}^*\mathbf{W}_o + \mathbf{W}_o\mathbf{A} + \mathbf{C}^*\mathbf{C} = 0. \tag{7.13b}$$

Lyapunov equations can be solved directly using the Bartels-Stewart method [11] under $O(n^3)$ runtime, which can be computationally intractable for high-dimensional systems with $n \gg 1$. For the systems considered here, it is still efficient to compute Gramians directly by solving Lyapunov equations, since state dimension in our examples is $n = 100$. We solve them using the Matlab control toolbox routine, `balreal`.

Fluid dynamics and climate simulations can easily reach state dimension $n = O(10^5)$ or higher due to fine grid discretization. Fortunately, balancing modes can also be efficiently computed using data-driven *empirical* Gramians. Here, the direct and adjoint modes are empirically computed from snapshots in time of impulse responses from the system and its adjoint. The products of these snapshot matrices closely approximate the true Gramians when the data sufficiently resolve the temporal dynamics and transients. Moore [113] was the first to realize that balancing transformations can be computed this way, and constructed the entire $n \times n$ gramians in this manner. More efficient strategies that leverage dimensionality reduction of snapshot matrices are currently used in practice. The first, developed by Willcox and Peraire [166], performs low-rank approximations of Gramians from direct and adjoint system snapshots. Shortly thereafter, Rowley [134] proposed computing modes using the product of direct and adjoint snapshot matrices, as well as a low-rank output projection of the adjoint system to dramatically reduce the number of requisite adjoint simulations.

7.3 Sensor and actuator placement using balanced truncation

We now develop a framework to identify sparse sensor and actuator locations using the balancing transformations defined above. Assuming a high-dimensional system is well-represented by its balanced truncation, the best states to observe and actuate are identified using partial observations and actuation in the truncated system. These sensor and actuator locations are determined using an efficient greedy QR pivoting algorithm. In particular, we optimize sensors assuming full actuation, and vice-versa, which allows us to select sensors and actuators independently of each other. We quantify the performance of the selected observations and actuation using a scalar measure of how well they can approximate the high-dimensional system, and derive bounds on this approximation error in terms of the Hankel singular values of the high-dimensional system.

7.3.1 Sensor placement

We begin with the balanced truncation (7.8) of a state-space system (7.1), and assume that its observable and controllable subspaces are well characterized by direct and adjoint modes $\boldsymbol{\Psi}_r$ and $\boldsymbol{\Phi}_r$. Partial observations are introduced using point sensors $\mathbf{C}$ in the balanced truncated system (7.8), so that $\hat{\mathbf{y}} = \mathbf{C}\boldsymbol{\Psi}_r\mathbf{a}$. As before, this corresponds to partial observations of the original system since $\hat{\mathbf{y}} \approx \mathbf{C}\mathbf{x}$. Then, sensor quality is quantified by how well the initial state $\mathbf{x}$ can be estimated from partial observations $\hat{\mathbf{y}}$ in balanced coordinates. The solution to this estimation problem is explicitly given by the following least-squares approximation

$$\hat{\mathbf{x}} = \boldsymbol{\Psi}_r(\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\mathbf{y} = \boldsymbol{\Psi}_r(\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\mathbf{C}\mathbf{x}, \tag{7.14}$$

which can be expressed as a projection $\mathbb{P}_C$ into the observable subspace

$$\mathbb{P}_C \triangleq \boldsymbol{\Psi}_r(\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\mathbf{C}. \tag{7.15}$$

Numerically, $\boldsymbol{\Psi}_r(\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\mathbf{C}$ is a suboptimal (non-orthogonal) projection onto the direct modes simply by definition of the biorthogonal balancing transformation, which corroborates the intuition of working with incomplete information (measurements $\mathbf{Cx}$ instead of $\mathbf{x}$). As in previous chapters, we pose the following numerical optimization problem for selecting point sensors $\mathbf{C}$ in terms of the approximation error between $\mathbf{x}$ and $\hat{\mathbf{x}}$, using the projection onto direct balancing modes $\mathbb{P}_C$

$$\min_{\mathbf{C}} \|\mathbf{x} - \boldsymbol{\Psi}_r(\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\mathbf{C}\mathbf{x}\|_2 \tag{7.16}$$

The remainder of this section details how to construct $\mathbf{C}$ and $\mathbf{B}$, and presents lower and upper bounds for the approximation (7.14) using our choice of $\mathbf{B}$ and $\mathbf{C}$. We propose a greedy construction of $\mathbf{C}$ using the matrix QR factorization with column pivoting of direct modes $\boldsymbol{\Psi}_r^*$, which expresses $\boldsymbol{\Psi}_r^*$ as the product of a unitary matrix $\mathbf{Q}$, an upper triangular matrix $\mathbf{R}$, and most importantly, the column permutation matrix $\mathbf{P}$

$$\boldsymbol{\Psi}_r^*\mathbf{P} = \mathbf{QR}. \tag{7.17}$$

As before, $\mathbf{C}$ is constructed from the first r columns of $\mathbf{P}$ transposed, and also given by Algorithm 2

$$\mathbf{C} \triangleq \mathbf{P}^T_{\cdot,j}, \text{ where } j : 1 \to r. \tag{7.18}$$

In other words, sensor selection corresponds to a rank-revealing row selection of the direct modes.

We use the fact that in the case of full observation, $\mathbf{C} = \mathbf{I}$ and the projection (7.14) is simply the approximation in balanced coordinates

$$\mathbf{x}_\star \triangleq \boldsymbol{\Psi}_r \boldsymbol{\Phi}_r^* \mathbf{x}, \tag{7.19}$$

in which case, the bounds on the approximation error are near-optimal.

Lemma 3 (Antoulas [6]). *The ℓ_2 error between the full state and its approximation using balanced coordinates is bounded by*

$$\|\mathbf{x} - \mathbf{x}_\star\|_2 \leq 2(\sigma_{r+1} + \cdots + \sigma_n). \tag{7.20}$$

An equivalent result holds for the adjoint system, for which the adjoint state projection onto balanced coordinates is given by $\mathbf{z}_\star \triangleq \boldsymbol{\Phi}_r \boldsymbol{\Psi}_r^* \mathbf{z}$, and

$$\|\mathbf{z} - \mathbf{z}_\star\|_2 \leq 2(\sigma_{r+1} + \cdots + \sigma_n). \tag{7.21}$$

Lemma 4 (Drmac & Gugercin [60]). *For any full-rank matrix $\boldsymbol{\Psi}_r \in \mathbb{R}^{n \times r}$ and its submatrix constructed from the first r rows of its column permutation matrix, $\mathbf{P}_r = \mathbf{P}_{\cdot,j}, j : 1 \to r$, obtained by QR factorization* (7.31)*, the spectral norm of $(\mathbf{P}_r^T \boldsymbol{\Psi}_r)^{-1}$ is bounded by*

$$\|(\mathbf{P}_r^T \boldsymbol{\Psi}_r)^{-1}\|_2 = \|(\boldsymbol{\Psi}_r^* \mathbf{P}_r)^{-1}\|_2 \leq \frac{\sqrt{n-r+1}}{\sigma_{\min}(\boldsymbol{\Psi}_r)} \frac{\sqrt{4^r + 6r - 1}}{3}, \tag{7.22}$$

where the bound grows as $\sqrt{n}O(2^r)$.

Theorem 1. *For any r-truncated row permutation matrix $\mathbf{C}$, the error from the projection* (7.14) *satisfies the following upper bound*

$$\|\mathbf{x} - \mathbb{P}_C \mathbf{x}\|_2 \leq \|\boldsymbol{\Psi}_r\|_2 \|(\mathbf{C}\boldsymbol{\Psi}_r)^{-1}\|_2 \|\mathbf{x} - \mathbf{x}_\star\|_2. \tag{7.23}$$

Proof. Define the residual between full state and its projection into balanced coordinates

$\mathbf{v} = \mathbf{x} - \mathbf{x}_\star$. Then

$$\begin{aligned}
\mathbb{P}_C\mathbf{v} &= \mathbb{P}_C\mathbf{x} - \mathbb{P}\mathbf{x}_\star \\
&= \mathbb{P}_C\mathbf{x} - \mathbf{\Psi}_r(\mathbf{C}\mathbf{\Psi}_r)^{-1}\mathbf{C}\mathbf{\Psi}_r\mathbf{\Phi}_r^*\mathbf{x}_\star \\
&= \mathbb{P}_C\mathbf{x} - \mathbf{\Psi}_r\mathbf{\Phi}_r^*\mathbf{x}_\star \\
&= \mathbb{P}_C\mathbf{x} - \mathbf{x}_\star,
\end{aligned}$$

where we use the fact that the orthogonal projection of $\mathbf{x}_\star$ onto $R(\mathbf{\Psi}_r)$ is $\mathbf{x}_\star$ again.

$$\begin{aligned}
\|\mathbf{x} - \mathbb{P}_C\mathbf{x}\|_2 &= \|(\mathbf{v} + \mathbf{x}_\star) - (\mathbb{P}_C\mathbf{v} + \mathbf{x}_\star)\|_2 \\
&= \|(\mathbf{I} - \mathbb{P}_C)\mathbf{v}\|_2 \\
&\leq \|\mathbb{P}_C\|_2\|\mathbf{x} - \mathbf{x}_\star\|_2 \\
&\leq \|\mathbf{\Psi}_r\|_2\|(\mathbf{C}\mathbf{\Psi}_r)^{-1}\|_2\|\mathbf{C}\|_2\|\mathbf{x} - \mathbf{x}_\star\|_2 \\
&= \|\mathbf{\Psi}_r\|_2\|(\mathbf{C}\mathbf{\Psi}_r)^{-1}\|_2\|\mathbf{x} - \mathbf{x}_\star\|_2.
\end{aligned}$$

□

Our logic here closely follows that in [41], which relies on $\mathbf{\Psi}_r$ restricted to be orthogonal ($\|\mathbf{\Psi}_r\|_2 = 1$). This property is not true in general for balanced modes.

Theorem 2. *The upper bound for the error between the full state and the projection is controlled by the sum of the discarded Hankel singular values and the condition number of direct modes $\kappa(\mathbf{\Psi}_r)$, where σ_i represents the ith Hankel singular value of the system*

$$\|\mathbf{x} - \mathbb{P}\mathbf{x}\|_2 \leq 2\alpha \sum_{i=r+1}^{n} \sigma_i, \tag{7.24}$$

$$\textit{where } \alpha = \kappa(\mathbf{\Psi}_r)\sqrt{n}O(2^r). \tag{7.25}$$

Proof. Combining Lemmas 3 and 4, with Theorem 1, we have

$$\begin{aligned}
\|\mathbf{x} - \mathbb{P}\mathbf{x}\|_2 &\le \|\mathbf{\Psi}_r\|_2 \|(\mathbf{C}\mathbf{\Psi}_r)^{-1}\|_2 \left(2 \sum_{i=r+1}^{n} \sigma_i \right) \\
&\le \|\mathbf{\Psi}_r\|_2 \left(\frac{\sqrt{n-r+1}}{\sigma_{\min}(\mathbf{\Psi}_r)} \frac{\sqrt{4^r + 6r - 1}}{3} \right) \left(2 \sum_{i=r+1}^{n} \sigma_i \right) \\
&= \kappa(\mathbf{\Psi}_r)\sqrt{n} O(2^r) \sum_{i=r+1}^{n} 2\sigma_i.
\end{aligned}$$

□

7.3.2 *Actuator placement*

Actuator placement is completely analogous to sensor placement, where we construct **B** from QR pivoting of the adjoint modes $\mathbf{\Phi}_r$ instead of the direct modes. In other words, we seek to place actuators along the most controllable directions in state space. Importantly, the controllability gramian in the original system is equivalent to the observability gramian of the adjoint system

$$\dot{\mathbf{z}} = \mathbf{A}^*\mathbf{z} + \mathbf{C}^*\mathbf{v} \tag{7.26a}$$

$$\mathbf{y} = \mathbf{B}^*\mathbf{z}, \tag{7.26b}$$

where we designate the output as **y** to be consistent with the previous discussion. As before, we desire the best actuators **B** in this adjoint system for estimating the adjoint state **z** from "measurements" (a slight abuse of terminology, since **B** corresponds to actuation in the original system), assuming there is no adjoint actuation ($\mathbf{v} = 0$). The adjoint system transforms using the same balanced coordinates defined above

$$\dot{\mathbf{b}}_r = \mathbf{\Psi}_r^*\mathbf{A}^*\mathbf{\Phi}_r\mathbf{b}_r \tag{7.27a}$$

$$\mathbf{y} = \mathbf{B}^*\mathbf{\Phi}_r\mathbf{b}_r. \tag{7.27b}$$

Similarly to the previous procedure, the adjoint state may be approximated in balanced coordinates using an analogous projection operator

$$\hat{\mathbf{z}} = \mathbf{\Phi}_r(\mathbf{B}^*\mathbf{\Phi}_r)^{-1}\mathbf{B}^*\mathbf{z} = \mathbb{P}_B\mathbf{z}. \tag{7.28}$$

Substituting $\mathbf{\Phi}_r$ for $\mathbf{\Psi}_r$, $\mathbf{z}$ for $\mathbf{x}$, and $\mathbf{B}^*$ for $\mathbf{C}$ in the proof for Theorem 1, we obtain the following corollary.

Corollary 1. *For any r-truncated row permutation matrix $\mathbf{B}^*$, the error from the projection* (7.28) *satisfies the following upper bound*

$$\|\mathbf{z} - \mathbb{P}_B\mathbf{z}\|_2 \leq \|\mathbf{\Phi}_r\|_2\|(\mathbf{B}^*\mathbf{\Phi}_r)^{-1}\|_2\|\mathbf{z} - \mathbf{z}_\star\|_2. \tag{7.29}$$

Thus the actuator placement problem reduces to controlling the growth of the term $\|(\mathbf{B}^*\mathbf{\Phi}_r)^{-1}\|_2$, which is computed using yet another QR factorization, this time of the adjoint modes

$$\mathbf{\Phi}_r^*\tilde{\mathbf{P}} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}. \tag{7.30}$$

Subsequently $\mathbf{B}$ is constructed from the first r columns of $\tilde{\mathbf{P}}$, given by Algorithm 2

$$\mathbf{B} \triangleq \tilde{\mathbf{P}}_{\cdot,j}, \text{ where } j : 1 \to r. \tag{7.31}$$

Corollary 2. *The upper bound for the error between the adjoint state and projection is controlled by the sum of the discarded Hankel singular values and the condition number of direct modes $\kappa(\mathbf{\Phi}_r)$, where σ_i represents the ith Hankel singular value of the system*

$$\|\mathbf{z} - \mathbb{P}_B\mathbf{z}\|_2 \leq 2\alpha \sum_{i=r+1}^{n} \sigma_i, \tag{7.32}$$

$$\textit{where } \alpha = \kappa(\mathbf{\Phi}_r)\sqrt{n}O(2^r). \tag{7.33}$$

Proof. This result immediately follows from the previous corollary and upon substituting $\mathbf{\Phi}_r$

for $\mathbf{\Psi}_r$ in Lemma 4. □

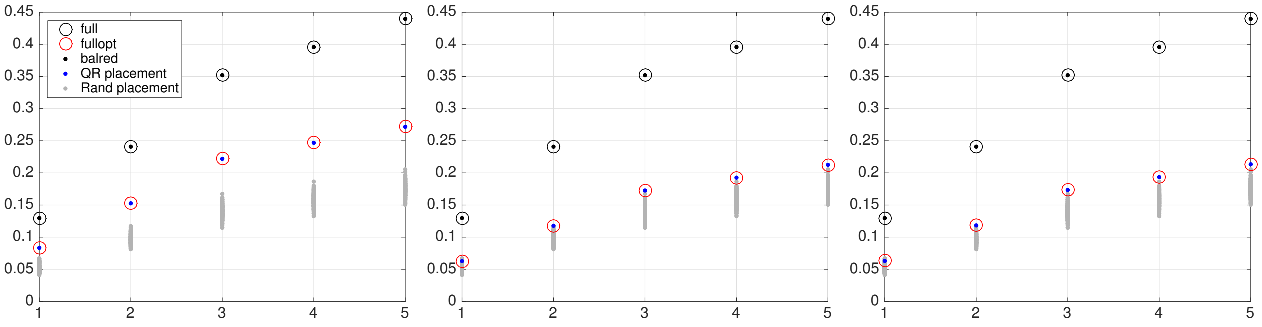
The above results are for the case when the number of sensors or actuators equal the rank of the truncated model ($p = q = r$), but they generalize to oversampling because the same upper bounds still hold for $p > r$ and $q > r$.

7.4 Results and discussion

This section demonstrates our sensor and actuator placement algorithm on two examples. The first is a random state-space model, for which we consider sensor placement and actuator placement independently, and then combined sensor and actuator placement. Hankel singular values are used to compare the input-output energy of the resulting systems with the original system. Next, we perform sensor and actuator placement on a linearized Ginzburg-Landau equation with stochastic disturbances, for which the H_2 optimal sensor and actuator placement has been determined by Chen and Rowley [43]. We compare the QR pivoting method with their gradient descent method for determining placements, and evaluate both sets of placements using their H_2 norm and LQG controller gain.

7.4.1 Random state space system

Our method is demonstrated on a 100 input, 100 output discrete random state-space model generated by the `drss()` command in MATLAB's control systems toolbox, and modify the command's output to have 100 point sensors and actuators for a fair comparison. That is, every state can be directly actuated and observed in the full system. The first three experiments optimize subsets of both sensors and actuators (fig. 7.2a), sensors only (fig. 7.2b), and actuators only (fig. 7.2c). We evaluate the quality of QR optimized configurations using each configuration's Hankel singular values, which quantify the the input-output energy of each configuration. Hankel singular values convey the degree of controllability and observability of the systems, and it is desired for the partially observed and actuated system's singular values to be as close as possible to that of the fully observed and actuated system.



(a) 40 QR sensors & actuators

(b) 40 QR sensors
Full actuation

(c) 40 QR actuators
Full observation

Figure 7.2: Sensor and actuator placement in random state-space system. The cumulative sum of Hankel singular values are plotted for various configurations, including 200 random ensembles of sensor or actuator placements (gray) in the balanced model truncated at ranks $r = 1$ to $r = 5$. The optimized QR configurations better preserve the input-output energy of the original system, whereas random sensors lag behind, especially when the system is both partially observed and partially actuated.

The cumulative sum of the normalized singular values are plotted in Figure 7.2; this provides a fair comparison the input-output energy retained in each system. Five sets of Hankel singular values are plotted of the following system variants: the full model, balanced model (`balreal`), full model with optimized placements (red), truncated model (`modred`), truncated model with QR optimized placements (blue), and finally, truncated model with random placements (gray). Our algorithm optimizes 40 sensors and/or actuators by truncated balanced modes at rank $r = 1$ to $r = 5$ on the horizontal axis. The more modes are retained, the better the chosen sensors and actuators characterize the input-output dynamics. Importantly, it can be seen that optimized placements on the balanced truncated system (blue dots) agree closely with optimized placements on the full model (red circles). This provides an empirical confirmation of our approach, which determines placements for the balanced truncated system assuming that they are optimal for the original system.

Because the system is randomly generated and dynamics are not sufficiently localized in state-space, many sensors and actuators (40% of state space) are required to characterize the system. In particular, this is reflected in its slowly decaying Hankel singular values. By contrast, the next example we consider is generated by a physical flow perturbation model, and has localized dynamics that allow for a more intuitive, visual interpretation of sensor and actuator placements.

7.4.2 Linearized Ginzburg-Landau with stochastic disturbances

The nonlinear Ginzburg-Landau equation models velocity perturbations in a given flow configuration, and in the case of small perturbations, the flow is well-described by the linearized equations. Consider the linearized Ginzburg-Landau system with stochastic process disturbance and noise ($\mathbf{d}$ and $\mathbf{n}$ respectively) present everywhere in the domain

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{D}^{1/2}\mathbf{d} \tag{7.34a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{N}^{1/2}\mathbf{n}, \tag{7.34b}$$

where $\mathbf{A}$ is the linearized Ginzburg-Landau operator discretized using Hermite pseudospectral differences

$$L \triangleq -\nu\frac{\partial}{\partial\xi} + \mu(\xi) + \beta\frac{\partial}{\partial\xi^2}. \tag{7.35}$$

Here ξ is the 1D spatial variable discretized at the roots of weighted Hermite polynomials, ν is the advection speed, β is the diffusion parameter, and $\mu(\xi)$ is the amplification factor. The H_2 optimal sensor and actuator placement for this system is determined by Chen and Rowley [43] using gradient minimization, and hence provides a benchmark comparison for our greedy placement. The H_2 norm of a system yields an averaged measure of output energy from impulsive inputs, and importantly, is invariant under similarity transformations. It can be computed from $\mathbf{B}, \mathbf{C}$, and the Gramians as follows

$$\|.\|_{H_2} = \text{trace}(\mathbf{C}\mathbf{W}_c\mathbf{C}^*) = \text{trace}(\mathbf{B}^*\mathbf{W}_o\mathbf{B}). \tag{7.36}$$

The observability and controllability gramians are not defined for this system since it

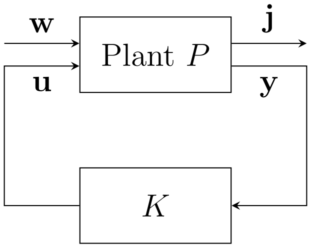


Figure 7.3: LQG controller for linearized Ginzburg-Landau

is unstable. We can, however, stabilize the system using linear quadratic Gaussian (LQG) control, which seeks to stabilize the state with minimal input using the cost function

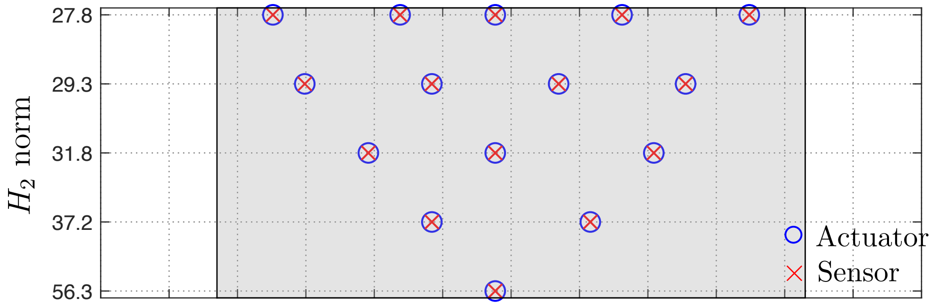
$$\mathbf{J} = \mathbf{x}^*\hat{\mathbf{Q}}\mathbf{x} + \mathbf{u}^*\hat{\mathbf{R}}\mathbf{u}, \tag{7.37}$$

where $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ refer to user-specified state and input weighting matrices (not to be confused with QR factorization). The structure of the controller is shown in Figure 7.3. Here $\mathbf{j}$ and $\mathbf{w}$ are concatenated vectors containing terms in the cost function, and the concatenated sources of process and measurement disturbance as defined below

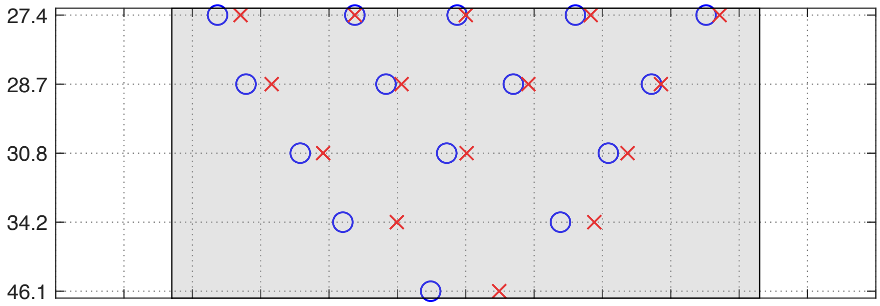
$$\mathbf{j} = \begin{bmatrix} \hat{\mathbf{Q}}^{1/2}\mathbf{x} \\ \hat{\mathbf{R}}^{1/2}\mathbf{u} \end{bmatrix}, \mathbf{w} = \begin{bmatrix} \mathbf{d} \\ \mathbf{n} \end{bmatrix}.$$

In essence, by minimizing $\mathbf{J}$, the controller minimizes the gain from the disturbances to the weighted cost vectors.

The resulting system is stable, therefore we can compute the balancing transformations using Matlab's `balreal()` command. Specifically, the new system matrices $\mathbf{A}_K, \mathbf{B}_K, \mathbf{C}_K$ are computed using Matlab's `lqg()` controller routine, and then passed as arguments to `balreal()` to compute balancing modes. We omit the details of the computation; explicit expressions for the controller system matrices are given in [43]. Sensing and actuation are defined on smooth spatial Gaussian kernels centered at the sensor/actuator location, and in



(a) Collocated QR pivoting placement

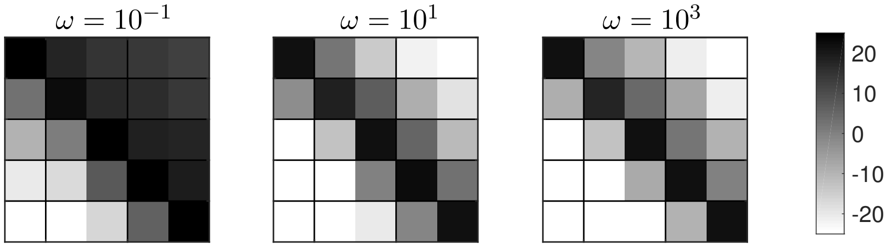


(b) QR initialized gradient minimization

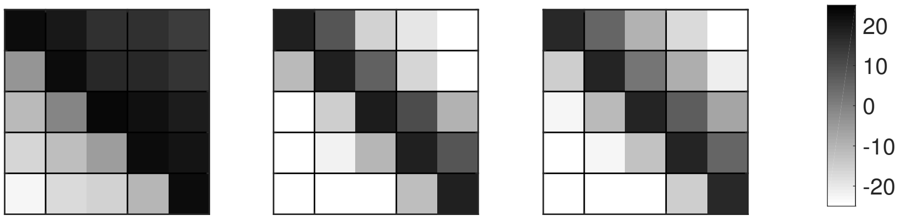
Figure 7.4: Sensor+actuator placement for linearized Ginzburg-Landau. Each row corresponds to the optimized placement for a certain budget of sensors and actuatorsPlacements based on QR pivoting of balanced truncated modes (a) closely approximate the H_2 norms of the placements determined using gradient minimization (b).

the small width limit of the kernels corresponds to point sensing and actuation. This permits derivative-based conjugate gradient minimization of the H_2 norm, as well as placement of sensors and actuators at locations that may not be gridpoints. The major drawback is that each Newton iteration requires $p+q$ solutions of $n \times n$ Lyapunov equations until convergence. Furthermore, the procedure requires optimizing an ensemble of random initial conditions to avoid converging to local minima. In [43], the optimal placement is computed using this conjugate gradient optimization for a spatial discretization corresponding to $n = 100$, and this method becomes computationally intractable as grid resolution increases.

Figure 7.4 compares the placement results from the QR algorithm and QR initialized gradient minimization. The resulting placements for $p = q = 1$ to $p = q = 5$ sensors and actuators are plotted from bottom to top, and the horizontal axis is the spatial domain extending from $\xi \in [-12, 12]$, with the wave amplification region shown in gray. For each selection of p, we truncate the balanced truncated modes to rank $r = p$, and pivot the truncated modes as described in the previous section. Notably, QR pivoting collocates sensors and actuators, and this agrees with a similar gradient optimization procedure that collocates placements for vibration control [80]. The corresponding H_2 norms of the resulting placement are displayed on the y-axis, from which it can be seen that the greedy QR placements closely approximate the optimal placements and as expected, land within the amplification region. In the five sensor and actuator case, the optimal placement [43, Figure 4] yields a H_2 norm of 27.4, which exactly agrees with our QR initialized placement and is approximated by the QR pivoted placement (27.8). Our results agree both visually and with H_2 norms with the optimal placements by Chen and Rowley. Yet another metric for evaluating the placement is the LQG gain of a given signal from a sensor to an actuator, and we compare the greedy QR placement to the QR initialized optimization using this metric over all sensors and actuators



(a) QR initialized gradient minimization



(b) QR pivoting

Figure 7.5: LQG gain in decibels for five sensor, five actuator system. Each block shows the gain from a signal $\exp(i\omega t)$ in sensor (column) k to actuator (row) j, ordered upstream to downstream.

in Figure 7.5. The results are indistinguishable from each other, and are also indistinguishable from the same LQG gain figure produced by Chen and Rowley [43, Figure 5]. Due to the dissipative nature of the system, it can be seen that low-frequency information from a sensor propagates to upstream actuators, but the majority of high-frequency information only to the nearest corresponding actuator.

QR pivoting runtime is only $O(nr^2)$ and the resulting placement's deviation from the optimal placements (fig. 7.4b) decreases with increasing p, q, and r. Furthermore, the conjugate gradient procedure initialized by the QR pivoted placement requires approximately 50 to 100 iterations to converge. Without QR initial placement, the conjugate gradient method requires an ensemble of random initial placements to converge.

We note that without balancing transformations, the Gramians are uninformative of where to place sensors. Chen and Rowley place a single sensor & actuator at the extrema of Gramian eigenvectors which yielded H_2 norms over 200, in comparison to their optimized results (and ours) which are roughly around 40. In contrast, we note that balanced truncated modes of the LQG controlled system are indeed useful for optimizing placements, and can be determined from a one-time computation of $O(n^3)$ to construct the LQG controller, and one more $O(n^3)$ computation to solve Lyapunov equations and compute the balancing modes.

7.5 Conclusions

In this chapter we develop a scalable sensor and actuator placement algorithm whose runtime scales only linearly with the number of state variables. Our approach relies on balanced model reduction [113, 134, 166], in which modes are hierarchically ordered by their observability and controllability. We extend EIM-based methods to sample the low-rank balancing modes of the system and determine maximally observable and controllable locations (sensor & actuators) in state space. The performance of this algorithm is demonstrated on a random state space system, and for controlling the linearized Ginzburg-Landau system with stochastic disturbances. Our optimized placements vastly exceed the performance of random placements as quantified by their Hankel singular value spectra, and closely approximate H_2

norms of optimal placements determined by expensive gradient minimization methods at a fraction of their runtime.

This work opens a variety of future directions in optimal QR-based sensor and actuator placement. So far our method relies on a known model of the dynamics, but it would also be interesting to generalize the method to data-driven system identification models. In addition, point sensors and actuators are simplifications of constrained or nonlinear sensing and actuation that may occur in practice. In these cases, one can generalize the QR pivoting algorithm to incorporate constraints [46], but nonlinear measurements or actuation remain an open challenge in optimal experiment design. Moreover, rapid advances in data collection yield extremely large search spaces, in which case the QR pivoting procedure may benefit from accelerations such as randomized or blocked pivoting.

Chapter 8

CONCLUSION

In this thesis we develop a data-driven, scalable sensor placement method using measurements derived from matrix pivots of low-dimensional embeddings, which are learned from high-dimensional data. These measurements are shown to be superior to competing measurement modalities (random, empirical interpolation and convex optimization) for global inference of complex systems. The compression of data under low-rank embeddings (such as POD and DMD) enables the reduction of high-dimensional signals with $O(10^4)$ and $O(10^6)$ points to sets of $10-100$ optimal sensors. We demonstrate this reduction on both simulated and observed high-dimensional data, for applications in image recovery, predictive manufacturing, ocean El Niño prediction, flow regime classification, and control. By exploiting known patterns in available data, we are able to significantly reduce the number of measurements required for signal recovery and global inference on complex static and dynamical systems.

Although we have developed pivoting-based sensing to be used in conjunction with data-driven POD, DMD or BPOD modes, the resulting method is quite general. Sensor placement via matrix pivoting admits any basis (any matrix) as input, and as long as one input dimension spans the ambient domain (usually space), it is extendable to nearly any linear inverse problem. Therefore, pivoting-based reconstruction is possible in more generic bases such as polynomials or radial basis functions. More generally, it may be used to select *experiments* or measurements in any candidate set of linear measurements, and thus can solve a wide class of problems originating in optimal experimental design, a connection made in Chapter 3.4. This generality lends itself to a wide variety of potential application scenarios, including 1) regimes where sensors are costly, and 2) regimes with massive data acquisition that would benefit from subsampling.

In regimes where sensors are expensive, the need for sensor optimization is clear. The number of sensors selected by these *rank-revealing* pivoting strategies is proportional to intrinsic rank, thus minimizing the number of sensors required for global inference once an appropriate low-rank representation is constructed. This minimization is extremely beneficial in physical systems such as ocean dynamics, fluid flows, robotics, and control (Chapters 3, 5–7). In addition, these applications may naturally impose rigid cost or location constraints on candidate locations, such as promoting measurements to lie along a wing or physical boundary of a fluid flow. Fortunately, pivoting algorithms easily admit the omission of certain measurement locations by simply removing columns of input, and can easily be adapted with weighted cost functions using user-specified weights on potential column pivots. This is an important avenue of ongoing work.

In contrast, high-fidelity data acquisition methods, including new and improved imaging technologies, are generating increasingly vast amounts of data. In such regimes, the computational burden is shifted from data collection to the immense storage and calculation capabilities required to process this data. Applications in imaging and predictive manufacturing immediately benefit from principled measurement strategies (Chapters 3–4). In the context of model reduction, interpolative measurement reduction combined with dimensionality reduction is known as *hyperreduction.* This is particularly important in applications where fast decisions are required, such as in high-performance computing or feedback control. For example, we have adapted pivoting to heavily downsample nonlinear, topologically-derived feature vectors for multi-way classification of human pose data in joint work with collaborators [75]. This represents an important step in generalizing hyperreduction to nonlinear observables of state space.

8.1 Extension to nonlinear dynamics and inference

Recent advances in dynamical systems are providing techniques to embed nonlinear systems in a linear framework through a suitable choice of *measurement functions* of the state, opening up the possibility of optimized sensing for nonlinear systems. As early as the 1930s, Koopman demonstrated that a nonlinear system can be rewritten as an infinite-dimensional linear operator on the Hilbert space of measurement functions [93]. This perspective did not gain traction until modern computation and data collection capabilities enabled the analysis of large volumes of measurement data.

Modern Koopman theory may drive sensor placement and the selection of nonlinear measurement functions on the sensors to embed nonlinear dynamics in a linear framework for optimal nonlinear estimation and control. This approach is consistent with neural control systems, where biological sensor networks (e.g., strain sensors on an insect wing) are processed through nonlinear neural filters before being used for feedback control. Much of the modern Koopman operator theory has been recently developed [111, 31, 32, 112], and it has been shown that under certain conditions DMD approximates the Koopman operator [137, 133, 157, 99]; sensor fusion is also possible in the Koopman framework [169]. Recently, Koopman analysis has been used to develop nonlinear estimators [143, 144] and controllers [26], although establishing rigorous connections to control theory is an ongoing effort [128, 130, 95]. Koopman theory has also been used to analyze chaotic dynamical systems from time-series data [69, 25], relying on the Takens embedding [146], which is related to sensor selection.

Beyond extending sensor selection to nonlinear systems and control, there is a significant opportunity to apply principled sensor selection to nonlinear inverse problems. The goal is to estimate latent variables from nonlinear measurements, and is typically solved in a probabilistic framework. Standard Bayesian approaches in data assimilation, such as the Kalman filter [88], estimate a probability distribution over the latent variable given the measurements, or the *posterior* distribution. Measurements can be chosen to minimize the posterior covariances, which is completely analogous to our method of choosing measurements to minimize error covariance. Likewise, it can be used for parameter selection in methods that propagate an ensemble of initial parameters samples forward such as the ensemble Kalman filter [64, 82] or other Bayesian inversion methods [141]. This is a promising avenue for future work, as optimal experimental design for nonlinear problems remains an open challenge.

BIBLIOGRAPHY

[1] NOAA optimal interpolation (OI) sea surface temperature (SST) v2. https://www.esrl.noaa.gov/psd/, July 2017.

[2] M. Abdo. *Theoretical and computational analysis of airfoils in steady and unsteady flows*. PhD thesis, McGill University, 2004.

[3] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A fast and scalable method for a-optimal design of experiments for infinite-dimensional bayesian nonlinear inverse problems. *SIAM Journal on Scientific Computing*, 38(1):A243–A272, 2016.

[4] A. A. Alonso, C. E. Frouzakis, and I. G. Kevrekidis. Optimal sensor placement for state reconstruction of distributed process systems. *AIChE Journal*, 50(7):1438–1452, 2004.

[5] J. C. Antolin-Urbaneja, J. Livinalli, M. Puerto, M. Liceaga, A. Rubio, A. San-Roman, and I. Goenaga. End-effector for automatic shimming of composites. Technical report, SAE Technical Paper, 2016.

[6] A. C. Antoulas. *Approximation of large-scale dynamical systems*. SIAM, 2005.

[7] T. Askham and J. N. Kutz. Variable projection methods for an optimized dynamic mode decomposition. *arXiv preprint arXiv:1704.02343*, 2017.

[8] R. G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118–120, 2007.

[9] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.

[10] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[11] R. H. Bartels and G. W. Stewart. Solution of the matrix equation AX+ XB= C. *Communications of the ACM*, 15(9):820–826, 1972.

[12] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

[13] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575, 1993.

[14] D. Bhattacharjee, M. Hemati, B. Klose, and G. Jacobs. Optimal actuator selection for airfoil separation control. In *2018 Flow Control Conference*, page 3692, 2018.

[15] J. M. Birch and M. H. Dickinson. Spanwise flow and the attachment of the leading-edge vortex on insect wings. *Nature*, 412(6848):729–733, 2001.

[16] D. A. Bistrian and I. M. Navon. Randomized dynamic mode decomposition for non-intrusive reduced order modelling. *International Journal for Numerical Methods in Engineering*, 2017.

[17] T. Bouwmans and E. H. Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.

[18] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[19] T. M. Boyl-Davis, D. D. Jones, and T. E. Zimmerman. Digitally designed shims for joining parts of an assembly, June 24 2014. US Patent 8,756,792.

[20] T. M. Boyl-Davis, D. D. Jones, and T. E. Zimmerman. Methods of fabricating shims for joining parts, August 30 2016. US Patent 9,429,935.

[21] I. Bright, G. Lin, and J. N. Kutz. Compressive sensing and machine learning strategies for characterizing the flow around a cylinder with limited pressure measurements. *Physics of Fluids*, 25(127102):1–15, 2013.

[22] R. E. Brown and M. R. Fedde. Airflow sensors in the avian wing. *Journal of Experimental Biology*, 179(1):13–30, 1993.

[23] B. W. Brunton, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Sparse sensor placement optimization for classification. *SIAM Journal on Applied Mathematics*, 76(5):2099–2122, 2016.

[24] B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of neuroscience methods*, 258:1–15, 2016.

[25] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz. Chaos as an intermittently forced linear system. *Nature Communications*, 8(19):1–9, 2017.

[26] S. L. Brunton, B. W. Brunton, J. L. Proctor, and J. N. Kutz. Koopman observable subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS ONE*, 11(2):e0150171, 2016.

[27] S. L. Brunton, J. L. Proctor, J. H. Tu, and J. N. Kutz. Compressed sensing and dynamic mode decomposition. *Journal of Computational Dynamics*, 2(2):165–191, 2015.

[28] S. L. Brunton, C. W. Rowley, and D. R. Williams. Reduced-order unsteady aerodynamic models at low Reynolds numbers. *Journal of Fluid Mechanics*, 724:203–233, 2013.

[29] S. L. Brunton, J. H. Tu, I. Bright, and J. N. Kutz. Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1716–1732, 2014.

[30] S. L. Brunton and J. N. Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge, 2019.

[31] M. Budišić and I. Mezić. An approximate parametrization of the ergodic partition using time averaged observables. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 3162–3168. IEEE, 2009.

[32] M. Budišić, R. Mohr, and I. Mezić. Applied Koopmanism a). *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.

[33] P. Businger and G. H. Golub. Linear least squares solutions by Householder transformations. *Numerische Mathematik*, 7(3):269–276, 1965.

[34] E. J. Candès. Compressive sensing. *Proceedings of the International Congress of Mathematics*, 2006.

[35] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[36] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications in Pure and Applied Mathematics*, 8(1207–1223), 59.

[37] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[38] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, pages 21–30, 2008.

[39] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011.

[40] W. F. Caselton and J. V. Zidek. Optimal monitoring network designs. *Statistics & Probability Letters*, 2(4):223–227, 1984.

[41] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[42] S. Chaturantabut and D. C. Sorensen. A state space error estimate for POD-DEIM nonlinear model reduction. *SIAM Journal on numerical analysis*, 50(1):46–63, 2012.

[43] K. K. Chen and C. W. Rowley. H_2 optimal actuator and sensor placement in the linearised complex Ginzburg-Landau system. *Journal of Fluid Mechanics*, 681:241–260, 2011.

[44] S. P. Chepuri and G. Leus. Continuous sensor placement. *IEEE Signal Processing Letters*, 22(5):544–548, 2015.

[45] B. Chouvion, A. Popov, S. Ratchev, C. Mason, and M. Summers. Interface management in wing-box assembly. Technical report, SAE Technical Paper, 2011.

[46] E. Clark, T. Askham, S. L. Brunton, and J. N. Kutz. Greedy sensor placement with cost constraints. *arXiv preprint arXiv:1805.03717*, 2018.

[47] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[48] E. S. Cole and J. Palka. The pattern of campaniform sensilla on the wing and haltere of drosophila melanogaster and several of its homeotic mutants. *Development*, 71(1):41–61, 1982.

[49] T. Collett and M. Land. Visual control of flight behaviour in the hoverfly *Syritta pipiens* L. *Journal of Comparative Physiology*, 99(1):1–66, 1975.

[50] T. Colonius and K. Taira. A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Computer Methods in Applied Mechanics and Engineering*, 197:2131–2146, 2008.

[51] S. A. Combes and T. L. Daniel. Into thin air: contributions of aerodynamic and inertial-elastic forces to wing bending in the hawkmoth *Manduca sexta*. *Journal of Experimental Biology*, 206(17):2999–3006, 2003.

[52] S. A. Combes. *Wing flexibility and design for animal flight*. PhD thesis, University of Washington, 2002.

[53] J. O. Dabiri. Optimal vortex formation as a unifying principle in biological propulsion. *Annual Review of Fluid Mechanics*, 41:17–33, 2009.

[54] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

[55] S. T. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3), 2016.

[56] L. Debnath. *Wavelet transforms and their applications*. Birkhäuser, 2002.

[57] B. H. Dickerson, Z. N. Aldworth, and T. L. Daniel. Control of moth flight posture is mediated by wing mechanosensory feedback. *The Journal of Experimental Biology*, 217(13):2301–2308, 2014.

[58] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[59] D. L. Donoho. 50 years of data science. In *Princeton NJ, Tukey Centennial Workshop*, 2015.

[60] Z. Drmac and S. Gugercin. A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions. *SIAM Journal on Scientific Computing*, 38(2):A631–A648, 2016.

[61] J. A. Duersch and M. Gu. True BLAS-3 performance QRCP using random sampling. *arXiv preprint arXiv:1509.06820*, 2015.

[62] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[63] N. B. Erichson, S. L. Brunton, and J. N. Kutz. Randomized dynamic mode decomposition. *arXiv preprint arXiv:1702.02912*, 2017.

[64] G. Evensen. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367, 2003.

[65] R. Everson and L. Sirovich. Karhunen–Loève procedure for gappy data. *Journal of the Optical Society of America A*, 12(8):1657–1664, 1995.

[66] M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.

[67] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.

[68] E. Gettrup. Sensory regulation of wing twisting in locusts. *Journal of Experimental Biology*, 44(1):1–16, 1966.

[69] D. Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *arXiv preprint arXiv:1507.02338*, 2015.

[70] A. C. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.

[71] A. C. Gilbert, J. Y. Park, and M. B. Wakin. Sketched SVD: Recovering spectral features from compressive measurements. *ArXiv e-prints*, 2012.

[72] G. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial & Applied Mathematics, Series B: Numerical Analysis*, 2(2):205–224, 1965.

[73] M. Grant, S. Boyd, and Y. Ye. CVX: Matlab software for disciplined convex programming, 2008.

[74] R. B. Guenther and J. W. Lee. *Partial differential equations of mathematical physics and integral equations*. Courier Corporation, 1988.

[75] W. Guo, K. Manohar, S. L. Brunton, and A. G. Banerjee. Sparse-TDA: Sparse realization of topological data analysis for multi-way classification. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2018.

[76] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[77] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta. De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets. *Theoretical and Computational Fluid Dynamics*, pages 1–20, 2017.

[78] M. S. Hemati, M. O. Williams, and C. W. Rowley. Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids*, 26(11), 2014.

[79] B. T. Hinson and K. A. Morgansen. Observability-based optimal sensor placement for flapping airfoil wake estimation. *Journal of Guidance, Control, and Dynamics*, 37(5):1477–1486, 2014.

[80] K. Hiramoto, H. Doki, and G. Obinata. Optimal sensor/actuator placement for active vibration control using explicit solution of algebraic riccati equation. *Journal of Sound and Vibration*, 229(5):1057–1075, 2000.

[81] P. J. Huber. John w. tukey's contributions to robust statistics. *Annals of statistics*, pages 1640–1648, 2002.

[82] M. A. Iglesias, K. J. Law, and A. M. Stuart. Ensemble kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 2013.

[83] J. Jamshidi, A. Kayani, P. Iravani, P. G. Maropoulos, and M. Summers. Manufacturing and assembly automation by integrated metrology systems for aircraft wing fabrication. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 224(1):25–36, 2010.

[84] S. Joshi and S. Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2009.

[85] M. R. Jovanović, P. J. Schmid, and J. W. Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2), 2014.

[86] J.-N. Juang and R. S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of guidance, control, and dynamics*, 8(5):620–627, 1985.

[87] J. Kahn and J. Tangorra. The effects of fluidic loading on underwater contact sensing with robotic fins and beams. *IEEE Transactions on Haptics*, 9(2), 2015.

[88] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[89] D. C. Kammer. Sensor placement for on-orbit modal identification and correlation of large space structures. *Journal of Guidance, Control, and Dynamics*, 14(2):251–259, 1991.

[90] J. Karhunen and J. Joutsensalo. Representation and separation of signals using nonlinear pca type learning. *Neural networks*, 7(1):113–127, 1994.

[91] C.-W. Ko, J. Lee, and M. Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43(4):684–691, 1995.

[92] D. Kondor, I. Csabai, L. Dobos, J. Szule, N. Barankai, T. Hanyecz, T. Sebok, Z. Kallus, and G. Vattay. Using robust pca to estimate regional characteristics of language use from geo-tagged twitter messages. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on*, pages 393–398. IEEE, 2013.

[93] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[94] B. Koopman and J. v. Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, 18(3):255–263, 1932.

[95] M. Korda and I. Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *arXiv preprint arXiv:1611.03537*, 2016.

[96] B. Kramer, P. Grover, P. Boufounos, S. Nabi, and M. Benosman. Sparse sensing and DMD-based identification of flow regimes and bifurcations in complex flows. *SIAM Journal on Applied Dynamical Systems*, 16(2):1164–1196, 2017.

[97] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(Feb):235–284, 2008.

[98] J. N. Kutz. *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*. Oxford University Press, 2013.

[99] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.

[100] J. N. Kutz, X. Fu, and S. L. Brunton. Multiresolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016.

[101] J. N. Kutz, X. Fu, S. L. Brunton, and N. B. Erichson. Multi-resolution dynamic mode decomposition for foreground/background separation and object tracking. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 921–929, Dec 2015.

[102] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 27(5):684–698, 2005.

[103] J. G. Leishman. *Principles of Helicopter Aerodynamics*. Cambridge University Press, 2006.

[104] D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.

[105] E. N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Project, Massachusetts Institute of Technology, Department of Meteorology, 1956.

[106] K. Manohar, S. L. Brunton, and J. N. Kutz. Environmental identification in flight using sparse approximation of wing strain. *Journal of Fluids and Structures*, 70:162–180, 2017.

[107] B. J. Marsh. Laser tracker assisted aircraft machining and assembly. Technical report, SAE Technical Paper, 2008.

[108] B. J. Marsh, T. Vanderwiel, K. VanScotter, and M. Thompson. Method for fitting part assemblies, July 13 2010. US Patent 7,756,321.

[109] P.-G. Martinsson. Blocked rank-revealing QR factorizations: How randomized sampling can be used to avoid single-vector pivoting. *arXiv preprint arXiv:1505.08115*, 2015.

[110] P.-G. Martinsson, G. Quintana Ortí, N. Heavner, and R. van de Geijn. Householder QR factorization with randomization for column pivoting (HQRRP). *SIAM Journal on Scientific Computing*, 39(2):C96–C115, 2017.

[111] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[112] I. Mezic. Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, 2013.

[113] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE transactions on automatic control*, 26(1):17–32, 1981.

[114] J. Muelaner, O. Martin, and P. Maropoulos. Achieving low cost and high quality aero structure assembly through integrated digital metrology systems. *Procedia CIRP*, 7:688–693, 2013.

[115] J. E. Muelaner, A. Kayani, O. Martin, and P. Maropoulos. Measurement assisted assembly and the roadmap to part-to-part assembly. In *7th international conference on digital enterprise technology*, pages 11–19. University of Bath, 2011.

[116] J. E. Muelaner and P. Maropoulos. Integrated dimensional variation management in the digital factory. In *7th International Conference on Digital Enterprise Technology*, pages 39–46. University of Bath, 2011.

[117] J. E. Muelaner and P. G. Maropoulos. Design for measurement assisted determinate assembly (MADA) of large composite structures. In *Journal of the Coordinate Metrology Systems Conference*. University of Bath, 2010.

[118] D. Needell and J. A. Tropp. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Communications of the ACM*, 53(12):93–100, 2010.

[119] T. Nestorović and M. Trajkov. Optimal actuator and sensor placement based on balanced reduced models. *Mechanical Systems and Signal Processing*, 36(2):271 – 289, 2013.

[120] B. R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, and F. Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, 2003.

[121] H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

[122] B. A. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004.

[123] L. Paninski. Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17(7):1480–1507, 2005.

[124] P. Y. Papalambros and D. J. Wilde. *Principles of optimal design: modeling and computation*. Cambridge university press, 2000.

[125] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(7–12):559–572, 1901.

[126] S. D. Pendergrass, J. N. Kutz, and S. L. Brunton. Streaming GPU singular value and dynamic mode decompositions. *arXiv preprint arXiv:1612.07875*, 2016.

[127] E. E. Postel and E. L. Leppert, Jr. Theoretical pressure distributions for a thin airfoil oscillating in incompressible flow. *Journal of the Aeronautical Sciences (Institute of the Aeronautical Sciences)*, 15(8), 1948.

[128] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[129] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[130] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Generalizing Koopman theory to allow for inputs and control. *arXiv preprint arXiv:1602.07647*, 2016.

[131] J. Ranieri, A. Chebira, and M. Vetterli. Near-optimal sensor placement for linear inverse problems. *IEEE Transactions on signal processing*, 62(5):1135–1146, 2014.

[132] D. K. Riskin, D. J. Willis, J. Iriarte-Díaz, T. L. Hedrick, M. Kostandov, J. Chen, D. H. Laidlaw, K. S. Breuer, and S. M. Swartz. Quantifying the complexity of bat wing kinematics. *Journal of Theoretical Biology*, 254(3):604–615, 2008.

[133] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. Henningson. Spectral analysis of nonlinear flows. *J. Fluid Mech.*, 645:115–127, 2009.

[134] C. W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15(03):997–1013, 2005.

[135] S. P. Sane, A. Dieudonné, M. A. Willis, and T. L. Daniel. Antennal mechanosensors mediate flight control in moths. *Science*, 315(5813):863–866, 2007.

[136] S. Sargsyan, S. L. Brunton, and J. N. Kutz. Nonlinear model reduction for dynamical systems using sparse sensor locations from learned libraries. *Physical Review E*, 92(033304), 2015.

[137] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, August 2010.

[138] P. Seshadri, A. Narayan, and S. Mahadevan. Effectively subsampled quadratures for least squares polynomial approximations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1003–1023, 2017.

[139] A. Sommariva and M. Vianello. Computing approximate Fekete points by QR factorizations of Vandermonde matrice. *Computers & Mathematics with Applications*, 57(8):1324–1336, 2009.

[140] S. Sterbing-D'Angelo, M. Chadha, C. Chiu, B. Falk, W. Xian, J. Barcelo, J. M. Zook, and C. F. Moss. Bat wing sensors support flight control. *Proceedings of the National Academy of Sciences*, 108(27):11291–11296, 2011.

[141] A. M. Stuart. Inverse problems: a bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[142] T. H. Summers, F. L. Cortesi, and J. Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Trans. Control of Network Systems*, 3(1):91–101, 2016.

[143] A. Surana. Koopman operator based observer synthesis for control-affine nonlinear systems. In *55th IEEE Conference on Decision and Control (CDC)*, pages 6492–6499. IEEE, 2016.

[144] A. Surana and A. Banaszuk. Linear observer synthesis for nonlinear systems using Koopman operator framework. *IFAC-PapersOnLine*, 49(18):716 – 723, 2016. 10th IFAC Symposium on Nonlinear Control Systems NOLCOS 2016.

[145] K. Taira and T. Colonius. The immersed boundary method: a projection approach. *Journal of Computational Physics*, 225(2):2118–2137, 2007.

[146] F. Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366–381, 1981.

[147] J. L. Tangorra, S. N. Davidson, I. W. Hunter, P. G. Madden, G. V. Lauder, H. Dong, M. Bozkurttas, and R. Mittal. The development of a biologically inspired propulsor for unmanned underwater vehicles. *Oceanic Engineering, IEEE Journal of*, 32(3):533–550, 2007.

[148] T. Theodorsen. General theory of aerodynamic instability and the mechanism of flutter. Technical Report 496, NACA, 1935.

[149] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[150] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[151] J. A. Tropp. Recovery of short, complex linear combinations via l_1 minimization. *IEEE Transactions on Information Theory*, 51(4):1568–1570, 2005.

[152] J. A. Tropp. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. *Signal Processing*, 86(3):589–602, 2006.

[153] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.

[154] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

[155] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[156] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond Nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Transactions on Information Theory*, 56(1):520–544, 2010.

[157] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, 2014.

[158] J. H. Tu, C. W. Rowley, J. N. Kutz, and J. K. Shang. Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. *Experiments in Fluids*, 55(9):1–13, 2014.

[159] D. Valenzuela, T. M. Boyl-Davis, and D. D. Jones. Systems, methods, and apparatus for automated predictive shimming for large structures, January 21 2015. US Patent App. 14/601,600.

[160] E. van den Berg and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, 2007.

[161] E. Van Den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[162] C. M. Vasquez, T. M. Boyl-Davis, D. I. Valenzuela, and D. D. Jones. Systems and methods for robotic measurement of parts, April 29 2014. US Patent App. 14/265,212.

[163] J. Videler, E. Stamhuis, and G. Povel. Leading-edge vortex lifts swifts. *Science*, 306(5703):1960–1962, 2004.

[164] W. J. Welch. Branch-and-bound search for experimental designs based on d optimality and other criteria. *Technometrics*, 24(1):41–48, 1982.

[165] K. Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & Fluids*, 35(2):208–226, 2006.

[166] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal*, 40(11):2323–2330, 2002.

[167] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[168] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis. A kernel-based method for data-driven Koopman spectral analysis. *Journal of Computational Dynamics*, 2(2):247–265, 2015.

[169] M. O. Williams, C. W. Rowley, I. Mezić, and I. G. Kevrekidis. Data fusion via intrinsic dynamic variables: An application of data-driven Koopman spectral analysis. *EPL (Europhysics Letters)*, 109(4):40007, 2015.

[170] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009.

[171] B. Yildirim, C. Chryssostomidis, and G. E. Karniadakis. Efficient sensor placement for ocean measurements using low-dimensional concepts. *Ocean Modelling*, 27(3):160–173, 2009.

[172] X. Yuan and J. Yang. Sparse and low-rank matrix decomposition via alternating direction methods. *preprint*, 12, 2009.

[173] R. Zimmermann and K. Willcox. An accelerated greedy missing point estimation procedure. *SIAM Journal on Scientific Computing*, 38(5):A2827–A2850, 2016.