Thermodynamic behavior of living systems: a biophysical approach to stochastic single-cell gene expression dynamics

Erin Angelini

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

Doctor of Philosophy

University of Washington

2024

Reading Committee:
Hong Qian, Chair
Ivana Bozic
Eric Shea-Brown

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

Thermodynamic behavior of living systems: a biophysical approach to stochastic single-cell gene expression dynamics

Erin Angelini

Chair of the Supervisory Committee:
Hong Qian
Applied Mathematics

Representing the state of a cell through its gene expression profile, we consider observable phenotypes as (metastable) attractor states of the underlying biochemical gene expression kinetics. The traditional deterministic dynamics, however, does not capture the possibility of spontaneous transitions between attractors (i.e., phenotype switching). In this work, we frame this picture in terms of stochastic dynamical systems and models: a quasi-potential "landscape" for an arbitrary dynamical system emerges as a large deviations rate function for the density of a singularly perturbed stochastic differential equation (SDE). This quasi-potential, which exists even for nonequilibrium biochemical dynamics in living cells, admits the most probable path between any two attractors and the characteristic time scale on which these transitions occur. We discuss the implications of this framework for the population dynamics at the level of cell cultures or tissues, specifically its applications to cancer population dynamics. We also consider two different frameworks for analyzing data from single-cell experiments through the lens of phenotypic attractors and state transitions. The first draws on the mathematics of thermodynamics, namely convex analysis and the Legendre-Fenchel transform, to derive an energy-like representation of an ergodic system from statistical measurements of the system. The latter is a maximum likelihood framework for inferring cell proliferation and phenotype switching rates from single-cell data, which we extend from a previous work to novel single-cell experiments using DNA-barcode lineage tracing. In addition to this work, we outline possible directions suitable for future research projects.

TABLE OF CONTENTS

Page

List of Figures iii

Glossary iv

Chapter 1: Introduction 1
1.1 Cellular heterogeneity and phenotype switching 1
1.2 Towards a rigorous "landscape theory" for cell phenotype 3
1.3 Quasi-potential function for non-gradient systems 7
1.4 Application to cancer development 9

Chapter 2: Quasi-Potential Function for Cell State Dynamics 11
2.1 Vector field decomposition for quasi-potentials 11
2.2 Large deviation principle for a quasi-potential 13
2.3 Stationary rate function as quasi-potential 19
2.4 Connection between rate function and entropy 21
2.5 Metastable states and the population balance equation 26

Chapter 3: Evolutionary Dynamics of Cancer Progression and Treatment 30
3.1 Phenotypic heterogeneity and "cancer attractors" 30
3.2 Mathematical models for cancer: mechanistic vs. descriptive 31
3.3 Drug resistance poses a major challenge to improving cancer treatment . . . 33
3.4 A dynamical model of drug-induced treatment resistance in cancer 36
3.5 Pharmacodynamic model of continuous therapy 40
3.6 Qualitative sensitivity analysis of pharmacodynamic parameters 42
3.7 Virtual patient cohort simulations 48
3.8 Discussion and conclusions 50

Chapter 4: Stochastic Thermodynamics for "Big Data" . . . 54
4.1 Gibbsian thermodynamics: from biophysical chemistry to living systems . . . 54
4.2 Statistical measurements, entropy, and the Legendre-Fenchel Transform . . . 58
4.3 Generalized Gibbs' theory . . . 64
4.4 Continuous state space and mechanical motion . . . 67
4.5 A theory of temperature and chemical potential of random motions . . . 74
4.6 Counting i.i.d. particles in $\mathbb{R}^3$ . . . 79
4.7 Discussion and conclusions . . . 82

Chapter 5: Quantifying Single-Cell Gene Expression Dynamics . . . 89
5.1 Direct estimation of cell state transition rates from single-cell data . . . 90
5.2 Demonstration of direct estimation method on *in silico* data . . . 94
5.3 Adapatation of a maximum likelihood framework for estimating cell state transition and growth rates . . . 97
5.4 Future directions . . . 106

Chapter 6: Conclusions and Future Directions . . . 109

Bibliography . . . 112

Appendix A: Quasi-Potential Theory . . . 126
A.1 Dominant balance for WKB approximation . . . 126
A.2 Asymptotic behavior of the Hessian matrix $\mathbf{H}_\varepsilon(\mathbf{x}_0)$ . . . 127

Appendix B: Cancer ODE Model . . . 129
B.1 Linear stability analysis . . . 129
B.2 Dynamics of tumor population recovery . . . 132
B.3 Supplemental tables and figures . . . 136

Appendix C: Stochastic Thermodynamics . . . 142
C.1 Tracking vs. counting and entropy of assimilation . . . 142
C.2 Lagrangian duality: multiplier as a conjugate variable . . . 144

LIST OF FIGURES

Figure Number Page

1.1 Waddington's epigenetic landscape . . . 2
1.2 From microscopic to macroscopic: parallels between protein conformations and cell state dynamics . . . 4
1.3 Example of a potential function for a gradient system . . . 6

3.1 Hanahan and Weinberg's hallmark traits of cancer cells . . . 31
3.2 Levels of parameterization in the dynamical model of tumor growth . . . 38
3.3 Coarse-grained sensitivity analysis of the effect of drug-induced resistance on treatment outcome . . . 43
3.4 Summary of model behavior for Cases **A** - **D** . . . 46
3.5 Analysis of time to progression over a wide range of values for potency and efficacy of inducing resistance . . . 48
3.6 Kaplan-Meier curves from virtual cohort simulations . . . 50

4.1 Entropy and free energy duality via the LFT. . . . 60
4.2 Derivative of cumulant generating function for a uniform point mass on the unit ball. . . . 63
4.3 Dual entropy-energy relations over a discrete state space ($n = 2$) . . . 73
4.4 Summary of relationship between classical mechanistic models and new stochastic thermodynamics . . . 87

5.1 Error of direct estimation method applied to *in silico* data . . . 96
5.2 Replication of confidence intervals from Gunnarsson *et al.* (2023) . . . 105

6.1 Idealized abstraction of a single cell as an ergodic, stochastic dynamical system 110

B.1 Dynamics of tumor population recovery for Case A . . . 138
B.2 Dynamics of tumor population recovery for Case B . . . 139
B.3 Dynamics of tumor population recovery for Case C . . . 140
B.4 Dynamics of tumor population recovery for Case D . . . 141

GLOSSARY OF TERMS AND NOTATION

We denote vectors by bold letters (e.g., $\mathbf{x}$); else, the object in question is a scalar (e.g., $\varphi(\mathbf{x})$) or a matrix (e.g., $D(\mathbf{x})$). We assume that vectors are column vectors. The only exception to this stylistic rule is the Hessian matrix of a function, which denote by $\mathbf{H}$. Otherwise, we use the following conventions.

$\mathbb{R}$ = the set of real numbers

$\mathbb{R}^m$ = m-dimensional Euclidean space

$\mathbb{R}^{m\times n}$ = the space of $m \times n$ matrices with real entries

$\mathbf{e}_j$ = the jth standard basis vector of $\mathbb{R}^m$ $(j = 1, 2, \dots m)$: $e_{j,i} = \delta_{ij}$

A^T ($\mathbf{x}^T$) = matrix (vector) transpose

$\langle \mathbf{x}, \mathbf{y} \rangle$ = inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$

$\nabla \cdot f(\mathbf{x})$ = divergence of the function $f : \mathbb{R}^m \to \mathbb{R}$

$\nabla \times \mathbf{F}(\mathbf{x})$ = curl of the function $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$

$\nabla^2 f(\mathbf{x})$ = Laplacian of the function $f : \mathbb{R}^m \to \mathbb{R}$

$\frac{\partial f(\mathbf{x})}{\partial \mathbf{v}}$ = directional derivative of the function $f : \mathbb{R}^m \to \mathbb{R}$ with respect to the vector $\mathbf{v}$

$f^*(\mathbf{y})$ = convex conjugate of the scalar function $f(\mathbf{x})$

$\sim$ = asymptotic equivalence relation OR distribution or a random variable

$\delta(\mathbf{x} - \mathbf{a})$ = Dirac delta function centered at $\mathbf{a} \in \mathbb{R}^m$

$\mathbf{W}_t$ = m-dimensional Weiner process with zero mean and unit variance

$\mathbb{E}[\mathbf{X}]$ = expectation of the random variable $\mathbf{X}$

a.s. = almost sure/almost surely

ACKNOWLEDGMENTS

I would like to thank my advisor, Hong Qian, and my *de facto* co-advisor, Sui Huang, for all the guidance, support, and many engaging discussions over the course of my PhD. I am also grateful for the input of my committee members, who have helped shape this dissertation over the past few months. I would also like to thank all the friends, family, loved ones, and various support groups who have guided me through the past six years in everything non-academic. Finally, I would like to thank my dog Penny, for whom I do everything.

Chapter 1

INTRODUCTION

1.1 Cellular heterogeneity and phenotype switching

The advent of single-cell technologies in the past few decades has put into sharp relief the need to understand the population dynamics of tissues and cell cultures at the level of the individual cell. In particular, having single-cell-level resolution in experimental data has driven the discovery of phenotypic variability in genetically identical, or *isogenic*, populations of cells [2]. This variability is often referred to as "cellular heterogeneity," or more specifically, "phenotypic heterogeneity." Such observations prompt the question of when this heterogeneity is biologically relevant, in that it serves a functional purpose for the population as a whole [2].

Phenotypic heterogeneity is a multifaceted property of a population of cells that can be caused by genetic factors, non-genetic factors, or both. Non-genetic phenomena include epigenetic alterations and directed changes in gene expression levels by a gene regulatory network (GRN) [62]. The study of these non-genetic behaviors (known more broadly as phenotype switching) and the dynamics underlying the phenotypic heterogeneity is fairly recent, leaving many open questions about the exact nature of phenotypic plasticity [62]. In particular, the ability of mature, differentiated cells to actively change their phenotypic state complicates the established notion of deterministic cell fate. Even the precise meaning of a "cell state" begs a more careful consideration.

One of the earliest such attempts at quantifying the factors that determine the developmental trajectory of a single cell was the "epigenetic landscape" metaphor proposed in the 1940s by Conrad Waddington (Figure 1.1) [135]. In this picture, a developing cell is a ball rolling around a physical landscape of differential gene expression; the valleys represent

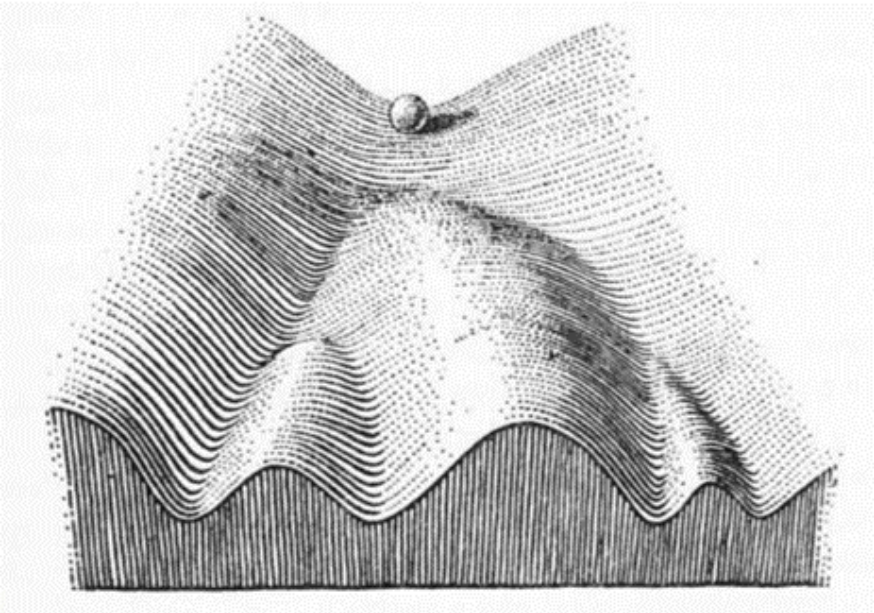

Figure 1.1: *Waddington's epigenetic landscape.* In this metaphor, the phenotypic development of a cell is represented by a ball rolling around in some predetermined landscape. The valleys, which represent the gene expression state, undergo bifurcations as the landscape evolves. The resulting branches represent the different possible cell fates. [135]

particular cell fates, and the hills represent phenotypic barriers between cell types (Figure 1.1) [135]. Although this picture is just a metaphor, the "ball-in-cup" heuristic is commonly used in the areas of statistical physics and thermodynamics as a way to represent the settling of a dissipative system into its stationary state [89]. From a deterministic dynamic perspective, the "ball" (i.e., the state of the system), rolls around a given landscape, going directly downhill until it reaches the bottom of a "cup" (i.e., a local minimum of the landscape) [89]. From a stochastic dynamic perspective, the ball undergoes continuous random perturbations to its position, allowing for movement upward on the landscape and even between "cups" on a long time scale [89]. Translating this heuristic, mental image into a rigorous mathematical theory for cellular development involves extending the equilibrium statistical physics to models of single cells in terms of stochastic mathematics.

1.2 Towards a rigorous "landscape theory" for cell phenotype

Our goal is to build a theoretical "landscape" framework to describe the time-dependent dynamics that drive cellular heterogeneity. In particular, we draw inspiration from the study of biophysical chemistry of single protein molecule dynamics [101]. In the latter theory, proteins are abstracted as systems of interacting atoms or (more coarse-grained) amino acid residues, and an *energy landscape* of molecular configurations emerges from the Newtonian dynamics arising from these interactions [33, 36, 92, 96, 103]. On the atomic level, the field of molecular dynamics (MD) uses the differential equations that govern Newtonian mechanics to simulate the interaction between the individual atoms of a single protein [96]. On the biopolymer level, energy landscape theory (ELT) treats a folding protein as a diffusing particle "searching" for energy-minimizing conformational states in a high-dimensional energy landscape [33, 36, 103]. Both of these theories are mechanistic, mathematical representations of single biological molecules in laboratories, which is our goal for the theory of cellular heterogeneity. In particular, the ELT framework of a particle diffusing on a landscape, and the underlying mathematics of its stochastic dynamics, lends itself to the representation of single *cells* (Figure 1.2).

Towards this end, we consider two distinct scales of phenotypic evolution: *microscopic* and *macroscopic*, and (Figure 1.2). We deliberately draw on the language of biophysical chemistry here because these physical scales of the state space tell us qualitatively about the amount of uncertainty inherent to each scale [101]. On the microscopic level, the "state" of a cell is the number of molecules of each reactant and product of the many biochemical reactions that drive the cell (Figure 1.2). The dynamics of this state follow some set of stochastic reaction-diffusion equations dictated by the chemical kinetics of each reaction. This scale is parallel to that of molecular dynamics for single proteins (Figure 1.2).

On the macroscopic level, these microscopic configurations produce some observable "state," e.g., the gene expression profile as measured the number of transcribed copies of each gene in some target set of genes (Figure 1.2). This state variable evolves according to

some stochastic dynamics (e.g., diffusion process), which approach a deterministic system in the many-molecule limit [101]. This gene expression profile also gives rise to discrete, observable cell phenotypes (§2.5) via the emergence of a *quasi-potential landscape* in this "small-noise" limit. This state variable is parallel to that of the observable conformationial state of a single protein (Figure 1.2).

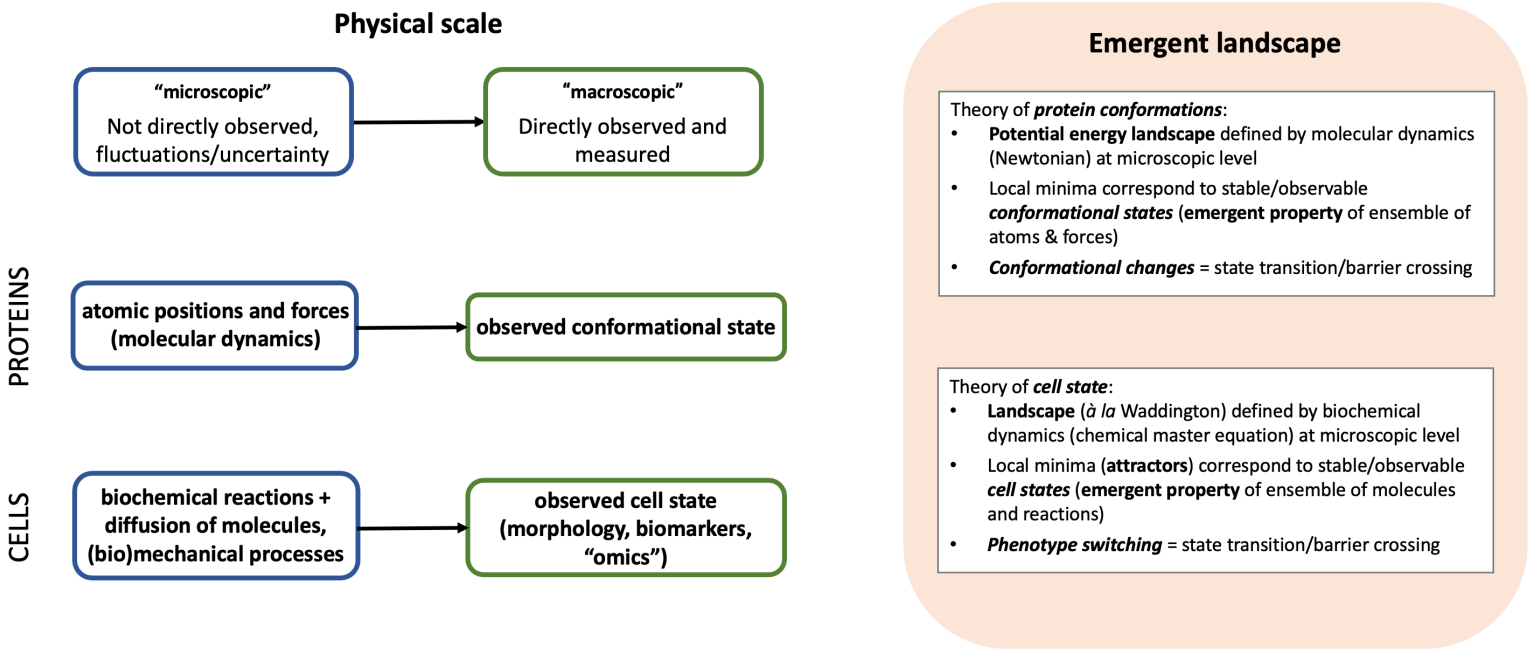

Figure 1.2: *From microscopic to macroscopic: parallels between protein conformations and cell state dynamics.* On the microscopic level, the cell "state" comprises of all the molecules contained within the cell and its membrane. These molecules, and the stochastic reaction-diffusion processes they undergo, direct the expression of a given set of genes. The resulting gene expression profile is given, for example, by the number of mRNA transcripts of each gene and evolves according to some stochastic dynamics. On the macroscopic level, the observable concentrations of these mRNA transcripts define the gene expression profile of the cell, which can be interpreted as the cell's phenotype. In this work, we discuss the quasi-potential landscape that emerges in the microscopic-to-macroscopic limit, akin to the well established theory of the conformational landscape of single protein molecules.

To frame this picture in terms of a mathematical theory, suppose we are concerned with the expression of a set of m particular genes. Then following the above description of the scales of gene expression, we represent the *macrostate* of a single cell by the vector of gene expression levels, given in mRNA concentration for each gene: $\mathbf{x}(t) \in \mathbb{R}^m$, where $x_i(t) =$ concentration of mRNA copies of gene i at time t [148]. As the GRN is a large, complex network of regulatory genes, we can assume that the time evolution of $\mathbf{x}(t)$ follows some highly nonlinear dynamical system $\frac{d\mathbf{x}}{dt} = b(\mathbf{x})$ [64, 65, 148]. A hallmark trait of many nonlinear dynamical systems is the existence of multiple attractors. In this framework, the attractors define the different possible phenotypes within the population of cells via their basins of attraction [148].

1.2.1 Potential function for gradient systems

In the case that we can write $\mathbf{b}$ in the form $\mathbf{b}(\mathbf{x}) = -\nabla U(\mathbf{x})$ for some scalar *potential* $U(\mathbf{x})$:

$$U(\mathbf{x}) = -\sum_{i=1}^{m} \int b_i(\mathbf{x}) dx_i \ .$$

The local extrema of $U(\mathbf{x})$ correspond exactly to the fixed points of the dynamical system, as $\mathbf{b}(\mathbf{x}) = 0$ iff $\nabla U(\mathbf{x}) = 0$. Furthermore, the potential function U is also a Lyapunov function of $\frac{d\mathbf{x}}{dt}$, as

$$\frac{d}{dt}\Big[U(\mathbf{x}(t))\Big] = -\nabla U(\mathbf{x}(t)) \cdot \nabla U(\mathbf{x}(t)) \leqslant 0 \ .$$

This Lyapunov property allows us to use U to characterize the asymptotic stability of the fixed points of the dynamics. In fact, it tells us that the stable fixed points correspond to the local minima of U, and the unstable fixed points sit at the local maxima or saddle points of U (Figure 1.3). Furthermore, the Lyapunov property is exactly what gives us the "ball-in-cup" heuristic - trajectories of the given dynamics decrease along the landscape $U(\mathbf{x})$ until settling at a local minimum. Having a gradient system means that the direction of this decrease is always in the direction of steepest descent along U.

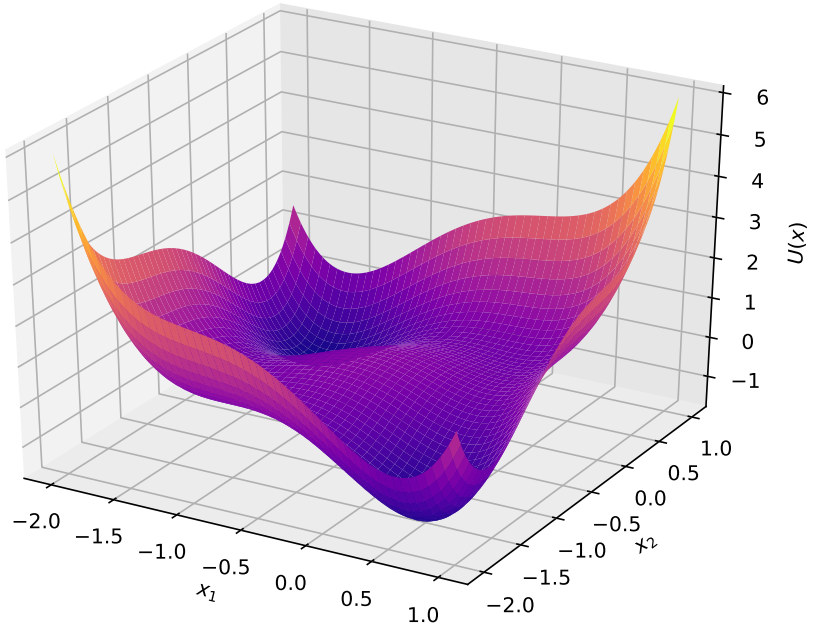

Figure 1.3: *Example of a potential function for a gradient system.* For $U(\mathbf{x}) = x_1^4 + x_2^4 + 2(x_1^3 + x_1^3) + x_1x_2 - \frac{x_1+x_2}{2}$, the corresponding gradient dynamical system $\frac{d\mathbf{x}}{dt} = -\nabla U(\mathbf{x})$ has four attractor states, which are located exactly at the base of the wells of $U(\mathbf{x})$.

In trying to understand the dynamics of phenotypic plasticity, it is insufficient to only consider the existence of attractors and their basins of attraction. This picture of stability does not allow for movement between attractors, which requires motions against a deterministic vector field. Instead, we must take a probabilistic approach to analyzing the gene expression dynamics: under continuous, random perturbation, the system can be "pushed" from one attractor state to another [38, 89]. Such an approach would allow us to characterize transitions between attractors in terms of the mean time scales and most likely path [38, 89].

Towards this end, consider small statistical fluctuations of our system within a given basin of attraction. We can think of many sources of this uncertainty: changes in the system's (i.e., the cell's) external environment, and the stochastic nature of biochemical reactions involving

fluctuating biopolymers.[1] Then the probability of the rare event that the system transitions from one basin of attraction to another is determined by the barrier height between each attractor and the saddle point that separates them along the least-action path [38, 43, 149]. That is, U not only determines the basin of attraction but also its depth relative to the surrounding saddle points. This property is exactly what allows one to use U to define the *relative stability* of two metastable states. In particular, we say that one attractor $\mathbf{x}_A^*$ is more stable than another $\mathbf{x}_B^*$ if the transition probability from $\mathbf{x}_A^*$ to $\mathbf{x}_B^*$ is greater than that from $\mathbf{x}_B^*$ to $\mathbf{x}_A^*$ [149].

1.3 Quasi-potential function for non-gradient systems

In general, the dynamics of a complex biological system do not admit a gradient system [134, 149]. Our goal is to find a function that replicates the "desired" properties of the potential function for a gradient system. Namely, for any dynamical system $\frac{d\mathbf{x}}{dt} = \mathbf{b}(\mathbf{x})$, we want to find a function $\varphi(\mathbf{x})$ such that

(i) φ is a *Lyapunov function* of the given dynamics;

(ii) φ admits *well defined* transition rates between two attractor states of the dynamics via the barrier height separating one attractor from another along φ.

From here on, we will refer to such a function as a *quasi-potential* for the given dynamics. These two conditions are the two properties of true potentials for gradient systems that we want to recover for any system. Condition (i) ensures that the quasi-potential is a "landscape" in the same sense as the true potential: it encodes the asymptotic stability of the system's attractors via the "ball-in-cup" heuristic. We will clarify what "well defined" means in condition (ii), but for now, it is sufficient to say that this requirement is nontrivial.

In order to understand condition (ii), we must consider the work of Freidlin and Wentzell. In their seminal work *Random Perturbations of Dynamical Systems,* they considered various

[1] In taking this approach, we are essentially assuming our system is "sufficiently small" (i.e., not macroscopic) so that observations of the system are subject to these environmental fluctuations.

large deviations principles for a singularly perturbed SDE on a finite time interval [38]:

$$d\mathbf{X}_t^{(\varepsilon)} = \mathbf{b}(\mathbf{X})dt + \varepsilon d\mathbf{W}_t \ .$$

This SDE is a *random perturbation* of the ODE $\frac{d\mathbf{x}}{dt} = \mathbf{b}(\mathbf{x})$. Given some initial condition $\mathbf{X}_{T_0}^{(\varepsilon)} = \mathbf{a}$, Freidlin and Wentzell consider the asymptotic behavior the probability that the process $\mathbf{X}_t^{(\varepsilon)}$ is "sufficiently close" to any smooth path $\mathbf{x}(t)$ for $t \in [T_0, T_1]$ as $\varepsilon \to 0$:

$$P\left(\left|\left|\mathbf{X}^{(\varepsilon)} - \mathbf{x}\right|\right| < \delta\right) \sim e^{-\frac{1}{\varepsilon^2} S_{T_0 T_1}(\mathbf{x})} \ .$$

This approach is often referred to as a *path integral formalism* for diffusion processes [38, 42, 135]. We can think of the functional S as the *action* along the path γ, where the most probable path of the process corresponds to the *least-action path* of S as $\varepsilon \to 0$ [38, 43]. Freidlin and Wentzell find that $S_{T_0 T_1}(\mathbf{x})$ takes the form

$$S_{T_0 T_1}(\mathbf{x}) = \frac{1}{2} \int_{T_0}^{T_1} \left| \frac{d\mathbf{x}}{dt} - \mathbf{b}\big(\mathbf{x}(t)\big) \right|^2 dt \ . \tag{1.1}$$

In the case that the deterministic dynamics $\frac{d\mathbf{x}}{dt}$ have multiple attractors, one can study the probability that the process $\mathbf{X}_t^{(\varepsilon)}$ escapes a given basin of attraction as the noise strength ε tends to 0. Using the least action functional, Freidlin and Wentzell construct a local quasi-potential by studying this exit problem [38, 43]. In particular, the *Freidlin-Wentzell quasi-potential* for two attractors $\mathbf{x}_A^*$ and $\mathbf{x}_B^*$ is defined as [38]:[2]

$$V(\mathbf{x}_A^*, \mathbf{x}_B^*) = \frac{1}{2} \inf \left\{ S_{T_A T_B}(\mathbf{x}) \mid \mathbf{x}(T_A) = \mathbf{x}_A^*, \quad \mathbf{x}(T_B) = \mathbf{x}_B^* \right\} \ . \tag{1.2}$$

This quasi potential gives the mean first exit time $\tau_{AB}^{(\varepsilon)}$ from the basin D_A of $\mathbf{x}_A^*$ into the

[2]Freidlin and Wentzell give a similar definition of the quasi-potential for limit cycles.

basin of $\mathbf{x}_B^*$ in the $\varepsilon \to 0$ limit [38]:

$$\tau_{AB}^{(\varepsilon)} \sim e^{\frac{1}{\varepsilon^2} \inf_{\mathbf{y} \in \partial D_A} V(\mathbf{x}_A^*, \mathbf{y})} \ . \tag{1.3}$$

In the $\varepsilon \to 0$ limit, Freidlin and Wentzell show that a continuous time Markov chain emerges between the attractors of $\frac{d\mathbf{x}}{dt}$ [38, 43]. Notably, the asymptotic behavior of the infinitesimal transition rate from one attractor state to another is given by [38]:

$$k_{AB} \equiv 1/\tau_{AB}^{(\varepsilon)} \sim e^{-\frac{1}{\varepsilon^2} \inf_{\mathbf{y} \in \partial D_A} V(\mathbf{x}_A^*, \mathbf{y})} \ . \tag{1.4}$$

We now have sufficient context to explain condition (ii) for a general quasi-potential: to admit "well defined" transition rates between two attractor states, a quasi potential must agree with the Freidlin-Wentzell quasi-potential along the least-action path. In Chapter 2 we discuss how to show this equivalence for a candidate quasi-potential. In particular, we present an approach that combines perturbation theory, the Fokker-Planck equation, and analytical mechanics to construct a quasi-potential that agrees with the Freidlin-Wentzell formulation. In §2.5 we discuss how this framework relates to the well-studied class of population balance models [101, 148].

1.4 Application to cancer development

The quasi-potential approach to describing cellular development has profound implications for the understanding of *oncogenesis*, or the process by which "normal" cells become malignant cancer cells. The classical paradigm of oncogenesis views this process as a linear, multi-step one in which normal cells acquire "hallmark" capabilities of cancer cells (e.g., sustained proliferation, angiogenesis, migration) via selection on a series of driver mutations [55, 56, 64]. By restricting its view to the genetic phenomena of Darwinian selection on isogenic clones, this framework neglects the phenotypic heterogeneity observed within cancer cell populations, as well as the effect of cell-cell interactions within this population on tumor

progression [55, 56, 64].

An alternate viewpoint to this genetic perspective of oncogenesis treats cancer as the result of incorrect regulation of the developmental pathways of a cell [64, 71]. This approach concerns itself not with the genetic makeup of a cell, but instead its gene expression profile, or *transcriptome* [64]. Cluster analysis of the transcriptomes of various tumors shows that cancer types fall into a small number of distinct groups, indicating that the gene expression profile provides a well-defined characterization of tumors [64]. This observation clarifies the notion of a "hallmark phenotype" of a cancer cell: such a phenotype is defined by a specific pattern of gene expression.

Defining the state variable of a single cell as its gene expression profile is exactly how the quasi-potential theory for cell development turns the metaphor of Waddington's landscape into a well-defined mathematical object. The similarity between these two concepts is no coincidence: the theory of "cancer attractors" uses the quasi-potential to describe oncogenesis as the perturbation of a cell's phenotype away from the "normal" developmental trajectory towards the attractor states that represent the hallmark traits of cancer cells [64, 65, 71]. In this framework, cancer is not simply the result a series of driver mutations, but one of many pre-existing cell fates [64, 65].

This theory of oncogenesis is particularly useful because the quasi-potential landscape encodes not only the genetic but also the non-genetic factors that contribute to differential gene expression. In particular, this theory explains the role of driver mutations in cancer progression through their effect on the landscape. Driver mutations effectively rewire the GRN, which then changes the gene expression dynamics $\frac{d\mathbf{x}}{dt} = b(\mathbf{x})$, which in turn alters the topography of the corresponding quasi-potential so that the probability of reaching the cancer attractor increases [64].

Chapter 2

QUASI-POTENTIAL FUNCTION FOR CELL STATE DYNAMICS

2.1 Vector field decomposition for quasi-potentials

Previous attempts at rigorously defining a quasi-potential for an arbitrary dynamical system $\frac{d\mathbf{x}}{dt} = \mathbf{b}(\mathbf{x})$ have focused on various decompositions of the vector field $\mathbf{b}(\mathbf{x})$. In particular, one looks to decompose $\mathbf{b}(\mathbf{x})$ into a gradient part and a remainder [135, 136, 149]:

$$\mathbf{b}(\mathbf{x}) = -\nabla\tilde{U}(\mathbf{x}) + \mathbf{F}_r\,. \tag{2.1}$$

Under the "right" decomposition, we get that $\tilde{U}(\mathbf{x})$ satisfies the two necessary quasi-potential conditions presented in §1.3. We will give a brief overview of various decompositions of the form (2.1) and the properties of the resulting $\tilde{U}$.

The first decomposition we consider is the *Helmholtz decomposition* for three-dimensional systems. This decomposition follows from the Helmholtz decomposition theorem, which tells us that we can write any smooth vector field as the sum of a gradient term and the curl of some vector field $\mathbf{A}$ [149]:

$$\mathbf{b}(\mathbf{x}) = -\nabla\tilde{U}^H(\mathbf{x}) + \nabla\times\mathbf{A}(\mathbf{x})\,.$$

If it indeed admits a quasi-potential, this decomposition is useful because it clearly breaks down the forces that drive movement of the dynamics along the landscape: the gradient term drives direct downhill motion between saddle points and attractors, and the curl term drives variation from these direct paths [149]. To find $\tilde{U}^H(\mathbf{x})$, we take the divergence of each side

of this equation, and use the fact that $\nabla \times \mathbf{A}(\mathbf{x})$ is divergence free:

$$\nabla \cdot \mathbf{b}(\mathbf{x}) = -\nabla \cdot \nabla \tilde{U}^H(\mathbf{x}) \equiv -\nabla^2 \tilde{U}^H(\mathbf{x}),$$

which is a *Poisson equation* for $\tilde{U}^H(\mathbf{x})$ [149]. Under appropriate boundary conditions, this equation will have a unique solution [149]. This solution, however, is not guaranteed to be a Lyapunov function for the dynamics $\frac{d\mathbf{x}}{dt} = \mathbf{b}(\mathbf{x})$. Therefore, this decomposition does not admit a quasi-potential.

Now, we will consider the approach taken by Wang *et al.*: a decomposition based on the stationary probability flux of a stochastic process centered about the deterministic dynamics [135, 136, 150]. If we consider the following vector valued Itô process driven by standard Brownian motion:

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X})dt + \sigma(\mathbf{X})d\mathbf{W}_t, \quad \boldsymbol{\sigma}(\mathbf{X}) \in \mathbb{R}^{m\times m},$$

the stationary density $f(\mathbf{x})$ of this process (if it exists) must satisfy the stationary Fokker-Planck equation:

$$0 = \nabla \cdot \left(\mathbf{b}(\mathbf{x})f(\mathbf{x}) + \nabla \cdot \left(D(\mathbf{x})f(\mathbf{x})\right)\right).$$

The matrix D of diffusion coefficients is defined as $2D(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{X})\boldsymbol{\sigma}(\mathbf{X})^T$. Define the *stationary probability flux* as

$$J(\mathbf{x}) = \mathbf{b}(\mathbf{x})f(\mathbf{x}) + \nabla \cdot \left(D(\mathbf{x})f(\mathbf{x})\right). \tag{2.2}$$

Taking $D = I$, we can decompose $\mathbf{b}(\mathbf{x})$ as follows [135, 136, 149, 150]:

$$\mathbf{b}(\mathbf{x}) = -\nabla \tilde{U}^{\text{prob}}(\mathbf{x}) + \frac{J(\mathbf{x})}{f(\mathbf{x})}, \quad \text{where} \quad \tilde{U}^{\text{prob}}(\mathbf{x}) = -\ln f(\mathbf{x}).$$

Unlike the Helmholtz decomposition, the remainder term $J(\mathbf{x})/f(\mathbf{x})$ is not necessarily divergence free [149]. Similar to the Helmholtz decomposition, however, the function $\tilde{U}^{\text{prob}}(\mathbf{x})$ is also not necessarily a Lyapunov function of the deterministic dynamics.

Finally, we will consider a decomposition based on orthognal vector fields. In particular, we look for $\tilde{U}^{\text{norm}}(\mathbf{x})$ and $\mathbf{F}_r(\mathbf{x})$ such that

$$\mathbf{b}(\mathbf{x}) = -\nabla \tilde{U}^{\text{norm}}(\mathbf{x}) + \mathbf{F}_r(\mathbf{x}) \ ,$$

where $\left\langle \nabla \tilde{U}^{\text{norm}}(\mathbf{x}), \mathbf{F}_r(\mathbf{x}) \right\rangle = 0$ [149]. In other words, we decompose $\mathbf{b}(\mathbf{x})$ into two orthogonal vector fields, one of which we can write as a gradient term. Unlike the previous decompositions, the function $\tilde{U}^{\text{norm}}(\mathbf{x})$ is a Lyapunov function of the dynamics, particularly because of this orthogonality condition.

In order for $\tilde{U}^{\text{norm}}$ to be a quasi-potential, it must satisfy the second necessary condition: agreement with the Freidlin-Wentzell quasi-potential along the least action path. Verifying this condition necessarily involves taking a probabilistic approach. Furthermore, unlike with the previous two decompositions, it is not immediately clear how to compute $\tilde{U}^{\text{norm}}$ given $\mathbf{b}(\mathbf{x})$. These questions motivate our large-deviations approach to finding the quasi-potential.

2.2 Large deviation principle for a quasi-potential

In this section, we lay out a derivation of a quasi-potential for an arbitrary dynamical system similar to Freidlin and Wentzell's. Specifically, we consider a singularly perturbed SDE and look for a large deviations rate function for the corresponding stationary probability density as the magnitude of the stochastic driving term tends to zero. Unlike Freidlin Wentzell theory, however, we derive an equation for this rate function from the stationary Fokker-Planck equation. From this equation, we can show that the rate function is indeed a quasi-potential for the deterministic dynamics in accordance with conditions (i) and (ii).

Consider the ODE for $\mathbf{x}(t) \in \mathbb{R}^m$, $t > 0$:

$$\frac{d\mathbf{x}}{dt} = \mathbf{b}(\mathbf{x}) \ , \tag{2.3}$$

and consider perturbing this ODE with a "small amount" of noise. That is, consider adding

to these dynamics $\sqrt{2\varepsilon}\,\boldsymbol{\sigma} d\mathbf{W}$, where $\boldsymbol{\sigma}\boldsymbol{\sigma}^T = D$ is an $m \times m$ positive definite matrix of diffusion coefficients:

$$d\mathbf{X} = \mathbf{b}(\mathbf{X})dt + \sqrt{2\varepsilon}\,\boldsymbol{\sigma}(\mathbf{X})\; d\mathbf{W}. \tag{2.4}$$

Let $\mathbf{X}_t^{(\varepsilon)}$ denote any stochastic process whose dynamics are governed by (2.4). We want to characterize the behavior of this process in the limit $\varepsilon \to 0$. In this sense, this singular perturbation of (2.3) represents going from a macroscopic system to a "small system," where the magnitude of noise ε scales with the inverse size of the system.

For a specified initial condition $\mathbf{X}_0^{(\varepsilon)} = \mathbf{a}$, let $f_\varepsilon(\mathbf{x}, t)$ denote the transition density of the stochastic process $\mathbf{X}_t^{(\varepsilon)}$ that solves (2.4):

$$f_\varepsilon(\mathbf{x}, t) = \frac{\partial^m}{\partial y_1 \dots \partial y_m} P\Big((\mathbf{X}_t^{(\varepsilon)})_i \leqslant y_i \;\forall\; i \;\Big|\; \mathbf{X}_0^{(\varepsilon)} = \mathbf{a}\Big)\Big|_{\mathbf{y}=\mathbf{x}}. \tag{2.5}$$

Then f_ε satisfies the *Fokker-Planck equation (FPE)* corresponding to SDE (2.4):

$$\frac{\partial f_\varepsilon}{\partial t} = -\nabla \cdot \left(\mathbf{b}(\mathbf{x}) f_\varepsilon(\mathbf{x}, t) - \varepsilon \sum_{j=1}^{m} \frac{\partial}{\partial x_j}\big[f_\varepsilon(\mathbf{x}, t) D_j(\mathbf{x})\big]\right), \quad \lim_{t\to 0} f_\varepsilon(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{a}), \tag{2.6}$$

where $D_j(\mathbf{x})$ is the j^{th} column of $D(\mathbf{x})$.

In the case of $\varepsilon = 0$, we want to establish a relation between $\mathbf{X}_t^{(\varepsilon)}$ and the dynamics given by (2.3). We will do so by considering the FPE. In this case, the FPE reduces to the following first-order PDE:

$$\frac{\partial f_0}{\partial t} = -\nabla \cdot \big(f_0(\mathbf{x}, t)\mathbf{b}(\mathbf{x})\big) = -f_0\big(\nabla \cdot \mathbf{b}(\mathbf{x})\big) - \mathbf{b}(\mathbf{x})^T \nabla f_0(\mathbf{x}, t), \tag{2.7}$$

which we can solve by the method of characteristics. The characteristic ODEs are indeed those given by the dynamics (2.3) subject to some initial condition:

$$\frac{d\mathbf{x}}{dt} = \mathbf{b}\big(\mathbf{x}(t)\big)\;, \quad \mathbf{x}(0) \equiv \mathbf{x}_0 \tag{2.8}$$

That is, the characteristic curves along which $f_0(\mathbf{x}, t)$ is transported satisfy (2.3).

Along these characteristics, f_0 satisfies the IVP

$$\frac{df_0}{dt} = -f_0\Big[\nabla \cdot \mathbf{b}\big(\mathbf{x}(t)\big)\Big], \quad \lim_{t\to 0} f_0(\mathbf{x}(t), t) = \delta(\mathbf{x}_0 - \mathbf{a})\,. \tag{2.9}$$

In practice, the initial condition $\mathbf{x}_0$ as a function of $\mathbf{x}$ and t is determined from the solution to (2.8), and is constant along the characteristic curves.[1] The general solution to this separable ODE for f_0 is given by

$$f_0(\mathbf{x}, t) = \delta\big(\mathbf{x}_0 - \mathbf{a}\big) e^{-\int_0^t \nabla \cdot b(\mathbf{x}(s;\ \mathbf{x}_0))ds}\,,$$

where $\mathbf{x}(t; x_0)$ solves (2.8) with initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Using the heuristic definition of the Dirac delta, we see that

$$f_0(\mathbf{x}, t) = \begin{cases} +\infty & \mathbf{x}_0 = \mathbf{a} \\ 0 & \mathbf{x}_0 \neq \mathbf{a} \end{cases}\,.$$

That is, the density function f_0 is a Dirac-delta distribution centered about the trajectory of the deterministic dynamics with $x_0 = \mathbf{a}$.

For $\varepsilon > 0$, a general solution to (2.6) is not readily available. Instead, we can draw on perturbation theory to determine the leading order behavior of its solutions. In particular, we want to find a rate function $\varphi(\mathbf{x}, t)$ such that[2]

$$\lim_{\varepsilon\to 0} \frac{f_\varepsilon(\mathbf{x}, t)}{e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x},t)}} = 1\,. \tag{2.10}$$

Equivalent to this limit definition, we write $f_\varepsilon(\mathbf{x}, t) \sim e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x},t)}$ as $\varepsilon \to 0$.

---

[1] This last statement is only true in the absence of shocks, i.e., intersecting characteristics.

[2] The relation '$\sim$' specifies neither a mode nor a domain of convergence, but we typically assume the "best case" scenario: uniform convergence on the entire domain of $f_\varepsilon(\mathbf{x}, t)$.

Taking the logarithm of both sides of (2.10), we get the equivalent statement

$$\lim_{\varepsilon\to 0} -\frac{1}{\varepsilon}\big(-\varepsilon \ln f_\varepsilon(\mathbf{x},t) - \varphi(\mathbf{x},t)\big) = 0\,. \tag{2.11}$$

This statement can only be true if $-\varepsilon \ln f_\varepsilon(\mathbf{x},t) - \varphi(\mathbf{x},t)\big)$ approaches zero on an faster scale than ε. From this statement, it follows that we must have that

$$\lim_{\varepsilon\to 0} -\varepsilon \ln f_\varepsilon(\mathbf{x},t) = \varphi(\mathbf{x},t) \tag{2.12}$$

for all $(\mathbf{x},t)$ such that (2.10) holds. In this sense, the asymptotic relation given in (2.10) defines a sort of *large deviations principle* for $f_\varepsilon(\mathbf{x},t)$ with rate function $\varphi(\mathbf{x},t)$. In general, we assume that $\varphi(\mathbf{x},t) \geqslant 0$ and has a global minimum for all $x \in \mathbb{R}^m, t > 0$.

2.2.1 Leading-order approximation via Fokker-Planck

For $0 < \varepsilon \ll 1$, we note that (2.6) is a singularly perturbed PDE. Thus, we can use a WKB approximation to derive an asymptotic representation of $f_\varepsilon(\mathbf{x},t)$ as $\varepsilon \to 0^+$:

$$f_\varepsilon(\mathbf{x},t) \sim \exp\left(\frac{1}{\delta}\sum_{n=0}^{\infty} \delta^n S_n(\mathbf{x},t)\right), \quad \delta = \varepsilon^q \text{ for some } q \in \mathbb{Q}. \tag{2.13}$$

In §A.1, we show that when (2.13) is substituted into the FPE, the dominant balance in ε is given by $\delta = \varepsilon$, and the leading order function S_0 satisfies

$$\frac{\partial S_0}{\partial t} = -\Big\langle \nabla S_0(\mathbf{x},t)\,,\, \mathbf{b}(\mathbf{x}) - D(\mathbf{x})\nabla S_0(\mathbf{x},t)\Big\rangle\,.$$

If we define $\varphi(\mathbf{x},t) = -S_0(\mathbf{x},t)$, the leading order asymptotic behavior of $f_\varepsilon(\mathbf{x},t)$ is given by

$$f_\varepsilon(\mathbf{x},t) \sim e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x},t)} \tag{2.14}$$

where $\varphi(\mathbf{x}, t)$ satisfies the PDE

$$-\frac{\partial \varphi}{\partial t} = \left\langle \nabla\varphi(\mathbf{x}, t)\,,\, \mathbf{b}(\mathbf{x}) + D(\mathbf{x})\nabla\varphi(\mathbf{x}, t) \right\rangle \tag{2.15}$$

with initial condition

$$\lim_{\varepsilon \to 0} e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x},0)} = \delta(\mathbf{x} - \mathbf{a})\,.$$

That is, we require that $e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x},0)}$ be some type of nascent delta function.

2.2.2 Stationary behavior

Towards the goal of quantifying the relative stability of the attractors of $\mathbf{b}(\mathbf{x})$, we want to find the *stationary* probability density function $f_\varepsilon(\mathbf{x})$ of the stochastic dynamics. If such a function exists, it must satisfy the stationary Fokker-Planck equation:

$$0 = \nabla \cdot \left(\mathbf{b}(\mathbf{x}) f_\varepsilon(\mathbf{x}) - \varepsilon \sum_{j=1}^{m} \frac{\partial}{\partial x_j} \left[f_\varepsilon(\mathbf{x}) D_j(\mathbf{x}) \right] \right). \tag{2.16}$$

We can think of finding the asymptotic behavior of this density as taking the $t \to \infty$ limit of $f_\varepsilon(\mathbf{x}, t)$, and then taking the $\varepsilon \to 0$ limit of the limiting function $f_\varepsilon(\mathbf{x})$. This order of limits, which represents two different time scales, is an important way in which the approach presented here differs from that of Freidlin and Wentzell. The Freidlin-Wentzell quasi-potential only deals with finite time processes in the limit $\varepsilon \to 0$, which essentially produces a *local* quasi-potential for each attractor [38, 150]. In contrast, finding the quasi-potential via the stationary distribution gives us *global* in that transition between multiple attractors is possible on a long time-scale [150]. We may encounter problems of existence, however, by assuming that the FPE has a stationary solution for $\varepsilon > 0$.

Again taking the leading order behavior of the WKB approximation, the stationary rate function φ,

$$\varphi(\mathbf{x}) = \lim_{\varepsilon \to 0} -\varepsilon \ln f_\varepsilon(\mathbf{x})\ , \tag{2.17}$$

satisfies the stationary PDE corresponding to (2.15):

$$0 = \Big\langle \nabla\varphi(\mathbf{x})\,,\, \mathbf{b}(\mathbf{x}) + D(\mathbf{x})\nabla\varphi(\mathbf{x}) \Big\rangle\,. \tag{2.18}$$

As with the time-dependent rate function, we require that $\varphi(\mathbf{x}) \geqslant 0$, and further require its global minimum to be $\varphi = 0$.

We note here that this rate function is essentially the leading order behavior of the quasi-potential proposed by Wang *et al.* [135, 136]. Therefore, although their quasi-potential is not a Lyapunov function of the deterministic dynamics, the *leading order* behavior φ is. We also note that when $D = I$, the above PDE tells us that we can write $\mathbf{b}(\mathbf{x})$ in accordance with Zhou *et al.*'s normal decomposition:

$$\mathbf{b}(\mathbf{x}) = \nabla\varphi(\mathbf{x}) + \mathbf{F}_r(\mathbf{x}), \quad \text{where} \quad \big\langle \nabla\varphi(\mathbf{x}), \mathbf{F}_r(\mathbf{x}) \big\rangle = 0\,.$$

Although we started with a probabilistic, large deviations approach to find a quasi-potential, the theory presented here relates to the vector field decomposition approach.

Returning to the equation (2.18), there are two degenerate solutions: $\nabla\varphi(\mathbf{x}) = 0$ and $\mathbf{b}(\mathbf{x}) + D(\mathbf{x})\nabla\varphi(\mathbf{x}) = 0$. The first case corresponds to $\varphi(\mathbf{x})$ being a constant function, which is not a "useful" quasi-potential in terms of defining relative stability of attractors. The second case we define as a *detailed balance* condition for the stationary rate function. Intuitively, drawing on the definition of detailed balance for Markov chains, this condition should give us a sense of reversibility for the stationary process. Specifically, it should imply a particular sense of "balance" in the probability flux.

Towards this end, we must first define the probability flux of a given state for a stochastic process. Consider writing the stationary FPE as $0 = \nabla \cdot J_\varepsilon(\mathbf{x})$, where

$$J_\varepsilon(\mathbf{x}) = \widetilde{\mathbf{b}}_\varepsilon(\mathbf{x}) f_\varepsilon(\mathbf{x}) - \varepsilon D(\mathbf{x})\nabla f_\varepsilon(\mathbf{x}), \quad \widetilde{\mathbf{b}}_\varepsilon(\mathbf{x}) = \mathbf{b}(\mathbf{x}) - \varepsilon \sum_{j=1}^{m} \frac{\partial}{\partial x_j}\big[D_j(\mathbf{x})\big]\,. \tag{2.19}$$

We call J_ε the *stationary probability flux*.[3] Substituting the asymptotic approximation $f_\varepsilon \sim e^{-\frac{1}{\varepsilon}\varphi}$ into $J(\mathbf{x})$, we get:

$$J_\varepsilon(\mathbf{x}) \sim \widetilde{\mathbf{b}}_\varepsilon(\mathbf{x}) e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x})} + D(\mathbf{x})\nabla\varphi(\mathbf{x}) e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x})}\,.$$

Applying the condition $D(\mathbf{x})\nabla\varphi(\mathbf{x}) = -\mathbf{b}(\mathbf{x})$, we obtain:

$$J_\varepsilon(\mathbf{x}) \sim \widetilde{\mathbf{b}}_\varepsilon(\mathbf{x}) e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x})} - \mathbf{b}(\mathbf{x}) e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x})} = -\varepsilon e^{-\frac{1}{\varepsilon}\varphi(\mathbf{x})} \sum_{j=1}^{m} \frac{\partial}{\partial x_j}\big[D_j(\mathbf{x})\big]\,.$$

Therefore, under the detailed balance condition, the stationary probability flux $J_\varepsilon(\mathbf{x})$ vanishes as $\varepsilon \to 0$.

2.3 Stationary rate function as quasi-potential

We now have the results necessary to prove that φ satisfies conditions (i) and (ii) for a quasi-potential of (2.3). First, we will show that it is a Lyapunov function of the dynamics. Namely, we will show that it is a positive definite function such that

$$\frac{d}{dt}\Big[\varphi\big(\mathbf{x}(t)\big)\Big] \leqslant 0,$$

with equality attained limit sets of (2.3).

Recall that, we require $\varphi(\mathbf{x}) \geqslant 0$, with equality only attained at the global minimum of φ. That is, we require φ to be a positive definite function *a priori*. Now, we will show that φ satisfies the conditions for asymptotic stability in the Lyapunov sense. Applying the chain rule along trajectories $\mathbf{x}(t)$ of (2.3),

$$\frac{d}{dt}\Big[\varphi\big(\mathbf{x}(t)\big)\Big] = \left\langle \nabla\varphi(\mathbf{x}), \frac{d\mathbf{x}}{dt} \right\rangle = \big\langle \nabla\varphi(\mathbf{x}), \mathbf{b}(\mathbf{x}) \big\rangle\,. \tag{2.20}$$

[3]This definition is consistent with the definition given in (2.2), except we now have the factor of the noise magnitude ε.

For all $\mathbf{x}(t) \neq \mathbf{x}^*$, where $\mathbf{b}(\mathbf{x}^*) = \mathbf{0}$, we can use (2.18) to write the above statement as follows:

$$\frac{d}{dt}\Big[\varphi\big(\mathbf{x}(t)\big)\Big] = -\Big\langle \nabla\varphi(\mathbf{x})\,,\, D(\mathbf{x})\nabla\varphi(\mathbf{x})\Big\rangle < 0, \tag{2.21}$$

as $D(\mathbf{x})$ is, by definition, positive definite. It follows from this statement that the equality $\frac{d}{dt}\Big[\varphi\big(\mathbf{x}(t)\big)\Big] = 0$ is attained on limit sets of (2.3). Therefore, $\varphi(\mathbf{x})$ satisfies the definition of a Lyapunov function for (2.3).

In the next section, we will see that $\varphi(\mathbf{x})$ also satisfies the second necessary condition to be a quasi-potential: under the coordinates $\mathbf{x}$ and momenta $\nabla\varphi(\mathbf{x})$, we can think of φ as the action of the system along the least-action path. In fact, this property directly relates our rate function φ to that proposed by Freidlin and Wentzell.

2.3.1 Least action principle and the Hamilton-Jacobi equation

For coordinates $\mathbf{q} = \mathbf{x}$ and momenta are given by $\mathbf{p} = \nabla\varphi(\mathbf{x}, t)$, we claim that the PDE (2.15) is a *Hamilton-Jacbobi*-type equation (*HJE*), with Hamiltonian

$$H(\mathbf{q}, \mathbf{p}) = \Big\langle \mathbf{p}\,,\, \mathbf{b}(\mathbf{q}) + D(\mathbf{q})\mathbf{p}\Big\rangle\,. \tag{2.22}$$

Similarly, the equation (2.18) for the stationary rate function $\varphi(\mathbf{x})$ is a *stationary* HJE. This property is what allows us to relate $\varphi(\mathbf{x})$ to the Freidlin-Wentzell quasi-potential along the least action path. Having this relation is exactly what allows us to use $\varphi(\mathbf{x})$ to compute the transition rates between two attractors.

To see exactly why this statement is true, recall that the Lagrangian $\mathcal{L}$ of a system is given by the *Legendre transform* of H with respect to $\mathbf{p}$ [139]:[4]

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) = \Big[\mathbf{p}\cdot\dot{\mathbf{q}} - H(\mathbf{q}, \mathbf{p})\Big]_{\mathbf{p}=\frac{1}{2}D^{-1}(\mathbf{q})(\dot{\mathbf{q}}-\mathbf{b}(\mathbf{q}))} = \frac{1}{4}\Big\langle D^{-1}(\mathbf{q})\big(\dot{\mathbf{q}} - \mathbf{b}(\mathbf{q})\big)\,,\, \dot{\mathbf{q}} - \mathbf{b}(\mathbf{q})\Big\rangle\,. \tag{2.23}$$

[4] We will discuss the Legendre transform further in §2.4.1 and §4.

From the Lagrangian, we define can define the action functional S:

$$S(\mathbf{q}, \dot{\mathbf{q}}) = \int_0^t \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) dt \equiv \frac{1}{4} \int_0^t \left\langle D^{-1}(\mathbf{q})\big(\dot{\mathbf{q}} - \mathbf{b}(\mathbf{q})\big)\,,\, \dot{\mathbf{q}} - \mathbf{b}(\mathbf{q}) \right\rangle dt\,. \tag{2.24}$$

In the derivation of the Hamilton-Jacobi equation, we require that this functional gives the *least action* by assuming that $\mathcal{L}$ satisfies the Euler-Lagrange equation [139]:

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}} = \frac{\partial \mathcal{L}}{\partial \mathbf{q}}\,. \tag{2.25}$$

This assumption is the "variational principle" mentioned at the beginning of the section. The HJE follows directly from this assumption:

$$-\frac{\partial S}{\partial t} = H\left(\mathbf{q}, \frac{\partial S}{\partial \mathbf{q}}\right)\,. \tag{2.26}$$

By taking $S = \varphi$ and $\mathbf{q} = \mathbf{x}$, we get back the PDE for the time dependent rate function (2.15). If we assume $\frac{\partial S}{\partial t} = \frac{\partial \varphi}{\partial t}$, we obtain the stationary PDE (2.18).

For coordinates $\mathbf{q} = \mathbf{x}$ and a finite time interval, the least action integral S defined in (2.24) is equal to twice the Freidlin-Wentzell action functional S_{0t} [38, 42, 89]. Therefore, if we solve (2.18) on a finite time domain with initial and final conditions at two given attractors, we can use the resulting solution to compute the transition rate from one attractor to another as outlined in §1.3 [149]. Although this problem can be solved (analytically or numerically) in low dimensions, existence of a solution is not generally guaranteed, especially in a high dimensional domain [42].

2.4 Connection between rate function and entropy

Not only does the connection between the rate function φ and the Hamiltonian (2.22) allow us to frame φ as a least-action functional, it also allows us to relate it to the *entropy* of the system. To see where this relation comes from, consider the equations of motion

corresponding to Hamiltonian (2.22):

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{b}(\mathbf{q}) + 2D(\mathbf{q})\mathbf{p} \tag{2.27}$$

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}} = -J_{\mathbf{b}}(\mathbf{q})^T\mathbf{p} - \sum_{i=1}^{m}\sum_{j=1}^{m} p_i p_j \frac{\partial}{\partial \mathbf{q}} D_{ij}(\mathbf{q}) \tag{2.28}$$

For this system, (2.27) tells us that $\frac{d\mathbf{x}}{dt}$ admits the deterministic dynamics at $\nabla\varphi(\mathbf{x},t) = 0$:

$$\frac{d\mathbf{x}}{dt}\Big|_{\nabla\varphi=0} \equiv \frac{d\mathbf{q}}{dt}\Big|_{\mathbf{p}=0} = \frac{\partial H}{\partial \mathbf{p}}\Big|_{\mathbf{p}=0} = \mathbf{b}(\mathbf{q}) \equiv \mathbf{b}(\mathbf{x}). \tag{2.29}$$

That is, if a path $\mathbf{x}(t)$ is such that the pair $(\mathbf{x}(t), t)$ admits a local extremum of $\varphi(\mathbf{x}, t)$, then it is a trajectory of the deterministic dynamics. Under the assumption that D is a positive definite matrix, the converse is true as well:

$$\frac{d\mathbf{x}}{dt} = \mathbf{b}(\mathbf{x}) \implies 2D(\mathbf{q})\nabla\varphi(\mathbf{x},t) = 0 \implies \nabla\varphi \equiv 0\,.$$

Therefore, $\mathbf{x}(t)$ is a trajectory of (2.3) if and only if the pair $(\mathbf{x}(t), t)$ is an extremum of the function $-\varphi(\mathbf{x}, t)$.

This conclusion highlights the natural association between large deviations theory and thermodynamics. In particular, if we think of the underlying process $\mathbf{X}_t^{(\varepsilon)}$ as a statistical observation of a macrostate of some m-dimensional system at time $t > 0$ with probability density f_ε. In the language of statistical mechanics, the entropy of a particular macrostate is proportional to the number of microstates that give rise to it. That is, the macrostates most likely to be observed are are ones with the highest entropy.

As we saw earlier in this section, for a given initial condition, the corresponding deterministic trajectory maximizes the probability distribution for $\varepsilon = 0$. Above, we found that trajectories $\mathbf{x}(t)$ of (2.3) admit the local extrema of $-\varphi(\mathbf{x}, t)$. It follows from the asymptotic form of the probability density $f_\varepsilon \sim e^{-\frac{1}{\varepsilon}\varphi}$ that the in the $\varepsilon \to 0$ limit, the maxima of $-\varphi$ correspond exactly to those of f_ε. The Lyapunov property of φ tells us that in the $t \to \infty$

limit, these local maxima are the attractors of (2.3). These results all indicate that $-\varphi$ is indeed larger for more probable macrostates, which means that it is a natural representation of the entropy of our system.

2.4.1 Entropy and free entropy as Legendre-Fenchel conjugates

Along these lines, we can think of the solution $-\varphi(\mathbf{x})$ of (2.18) as the stationary entropy of a macrostate $\mathbf{x}$. This thermodynamic approach to the large deviations principle for f_ε allows us to define an additional stationary quantity: the *free entropy* of the system. More specifically, from the stationary density $f_\varepsilon(\mathbf{x})$, define the *partition function* $Z_\varepsilon(\boldsymbol{\beta})$:

$$Z_\varepsilon(\boldsymbol{\beta}) \equiv \int_{\mathbb{R}^m} f_\varepsilon(\mathbf{x}) e^{-\frac{1}{\varepsilon}\boldsymbol{\beta}\cdot\mathbf{x}} d\mathbf{x} \,, \quad \boldsymbol{\beta} \in \mathbb{R}^m \,. \tag{2.30}$$

When $m = 1$, we can think of the parameter β as the thermodynamic quantity *inverse temperature* [82]. For $m > 1$, each parameter β_i typically represents an intensive quantity divided by temperature [82]. We note that $Z_\varepsilon(\boldsymbol{\beta}) = M_\varepsilon(-\boldsymbol{\beta}/\varepsilon)$, where $M_\varepsilon(\mathbf{s})$ denotes the joint moment generating function associated with f_ε.

From this partition function, we can define the (Massieu) *free entropy* of our system $\mathcal{A}(\beta)$ through the following limit:

$$\mathcal{A}(\boldsymbol{\beta}) = \lim_{\varepsilon \to 0} -\varepsilon \ln Z_\varepsilon(\boldsymbol{\beta}) \,. \tag{2.31}$$

As we did with the limit (2.17) for $\varphi(\mathbf{x})$, we assume that this limit exists and is uniform for all $\boldsymbol{\beta}$ in some subset of $\mathbb{R}^m$.

Under the assumption that the convergence in (2.17) is uniform, along with an additional assumption on the extrema of $-\varepsilon \ln f_\varepsilon(\mathbf{x})$, we can show that the above limit is equivalent to the *Legendre-Fenchel transform (LFT)* of $-\varphi(\mathbf{x})$:

$$\mathcal{A}(\boldsymbol{\beta}) = \inf_{\mathbf{x} \in \mathbb{R}^m} \left\{ \langle \boldsymbol{\beta}, \mathbf{x} \rangle + \varphi(\mathbf{x}) \right\} . \tag{2.32}$$

Keeping with our thermodynamic framing, we state this result as follows: the free energy

is the *convex conjugate* of the negative entropy, with $\boldsymbol{\beta}$ being the *conjugate variable* to the macrostate $\mathbf{x}$.

We will use Laplace's method of estimating integrals to derive (2.32), as the approximation derived from this method becomes exact in the limit of $n \to \infty$. We first note that we can write $Z_\varepsilon(\boldsymbol{\beta})$ as follows:

$$Z_\varepsilon(\boldsymbol{\beta}) = \int_{\mathbb{R}^m} e^{-\frac{1}{\varepsilon}\langle \boldsymbol{\beta}, \mathbf{x}\rangle + \ln f_\varepsilon(\mathbf{x})} d\mathbf{x} = \int_{\mathbb{R}^m} e^{-\frac{1}{\varepsilon}\left(\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})\right)} d\mathbf{x} \ ,$$

If we assume that $\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})$ has a global minimum at some point $\mathbf{x} = \mathbf{x}_0$, we can apply Laplace's method to the above integral to get its leading order approximation as $\varepsilon \to 0$:

$$Z_\varepsilon(\beta) \sim \frac{(2\pi\varepsilon)^{m/2}}{\sqrt{\det \mathbf{H}_\varepsilon(\mathbf{x}_0)}} e^{-\frac{1}{\varepsilon} \inf_{\mathbf{x}\in\mathbb{R}^m}\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})\}} \ ,$$

where $\mathbf{H}_\varepsilon(\mathbf{x}_0)$ denotes the Hessian matrix of $-\varepsilon \ln f_\varepsilon(\mathbf{x})$ evaluated at $\mathbf{x} = \mathbf{x}_0$. Then to leading order in ε, the logarithm of $Z_\varepsilon(\boldsymbol{\beta})$ behaves according to

$$\ln Z_\varepsilon(\boldsymbol{\beta}) \sim -\frac{1}{\varepsilon} \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})\right\} - \frac{1}{2} \ln \det \mathbf{H}_\varepsilon(\mathbf{x}_0) + o(\varepsilon^{-1}).$$

Under appropriate conditions, we can further show that[5] as $\varepsilon \to 0$,

$$\ln \det \mathbf{H}_\varepsilon(\mathbf{x}_0) = o(\varepsilon^{-1})$$

which tells us that

$$-\varepsilon \ln Z_\varepsilon(\boldsymbol{\beta}) \sim \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})\right\} + \varepsilon o(\varepsilon^{-1}). \tag{2.33}$$

By definition of little-o notation, $\varepsilon o(\varepsilon^{-1}) \to 0$ as $\varepsilon \to 0$. Therefore, in taking $\varepsilon \to 0$ on both

[5] See Appendix A.2 for full derivation.

sides of (2.33), the asymptotic approximation becomes exact, and we have:

$$\lim_{\varepsilon\to 0} -\varepsilon \ln Z_\varepsilon(\boldsymbol{\beta}) = \lim_{\varepsilon\to 0} \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})\right\}$$

With the assumption that the convergence of (2.17) is uniform, we can interchange the $\varepsilon \to 0$ limit with taking the infimum over $x \in \mathbb{R}^m$:

$$\lim_{\varepsilon\to 0} \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \varepsilon \ln f_\varepsilon(\mathbf{x})\right\} = \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle - \lim_{\varepsilon\to 0} \varepsilon \ln f_\varepsilon(\mathbf{x})\right\} = \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle + \varphi(\mathbf{x})\right\},$$

Therefore, we indeed have that

$$\lim_{\varepsilon\to 0} -\varepsilon \ln Z_\varepsilon(\boldsymbol{\beta}) \equiv \mathcal{A}(\beta) = \inf_{\mathbf{x}\in\mathbb{R}^m} \left\{\langle \boldsymbol{\beta}, \mathbf{x}\rangle + \varphi(\mathbf{x})\right\}.$$

□

If we assume that $\varphi(\mathbf{x})$ is convex (i.e., its Hessian is positive definite), then the LFT is an involution. That is, if we know $\mathcal{A}(\boldsymbol{\beta})$, we can recover $-\varphi(\mathbf{x})$ via the transform $\mathcal{A}^*$ of $\mathcal{A}$:

$$-\varphi(\mathbf{x}) = \sup_{\beta\in\mathbb{R}^m} \left\{\langle \mathbf{x}, \boldsymbol{\beta}\rangle - \mathcal{A}(\boldsymbol{\beta})\right\}. \tag{2.34}$$

In this manner, the LFT defines an invertible relation between free energy $\mathcal{A}(\boldsymbol{\beta})$ and entropy $-\varphi(\mathbf{x})$. Otherwise, $\mathcal{A}^*(\mathbf{x})$ is the *convex hull* of $-\varphi(\mathbf{x})$.

In addition to giving us a relation between the conjugate functions $\mathcal{A}$ and φ^*, the LFT also tells us how the conjugate variables $\mathbf{x}$ and $\boldsymbol{\beta}$ depend on each other. In addition to convexity, further assume that $\varphi(\mathbf{x})$ and $\mathcal{A}(\boldsymbol{\beta})$ are differentiable on their respective domains. Then computing their LFT reduces to finding where the derivative of the argument is equal to zero. That is,

$$\mathcal{A}(\boldsymbol{\beta}) = \langle \boldsymbol{\beta}, \mathbf{x}\rangle + \varphi(\mathbf{x}) \iff \langle \boldsymbol{\beta}, \mathbf{x}\rangle + \varphi(\mathbf{x}) = \inf_{\boldsymbol{\xi}\in\mathbb{R}^m} \{\langle \boldsymbol{\beta}, \boldsymbol{\xi}\rangle + \varphi(\boldsymbol{\xi})\} \iff \boldsymbol{\beta} = -\nabla\varphi(\mathbf{x}).$$

Similarly, we have that

$$-\varphi(\mathbf{x}) = \langle \mathbf{x}, \boldsymbol{\beta} \rangle - \mathcal{A}(\boldsymbol{\beta}) \iff \langle \mathbf{x}, \boldsymbol{\beta} \rangle - \mathcal{A}(\boldsymbol{\beta}) = \sup_{\boldsymbol{\eta} \in V} \{ \langle \mathbf{x}, \boldsymbol{\eta} \rangle - \mathcal{A}(\boldsymbol{\eta}) \} \iff \mathbf{x} = \nabla \mathcal{A}(\boldsymbol{\beta}) \, .$$

In sum, x and β satisfy the following dual relationship:

$$\mathbf{x}(\boldsymbol{\beta}) = \nabla \mathcal{A}(\boldsymbol{\beta}), \qquad \boldsymbol{\beta}(\mathbf{x}) = -\nabla \varphi(\mathbf{x}) \ , \tag{2.35}$$

which tells us that $\mathcal{A}$ and $-\varphi$ are both *thermodynamic potentials*. This dual relation further establishes a connection between the rate function φ and the language of thermodynamics. As we will discuss in Chapter 4, it also admits an energetic representation of the underlying system when the relative frequency over the system's states is estimated from an infinite number of data points.

2.5 Metastable states and the population balance equation

Returning to our broader goal of using this quasi potential framework to characterize gene expression and cell fates, again consider the case where $x_i(t)$ denotes the concentration of mRNA copies of gene i at time t, with dynamics $\frac{d\mathbf{x}}{dt} = b(\mathbf{x})$. Each possible individual cell *type* is defined as a (meta)stable steady-state of these dynamics: the states $\mathbf{x}(t)$ that lie within the basin of attraction of some asymptotically stable fixed point $\mathbf{x}^*$. These basins are separated by barriers as defined by the Freidlin-Wentzell quasi potential in (1.2), which is in theory attainable from the dynamics via the rate function φ, as laid out in §2.2. This general definition of cell state encapsulates both the concrete idea of a phenotype, as well as the more loosely defined notion of a "subtype" within a multicellular population.

At a higher level, a population of cells with the identical genomes (*isogenic*) can consist of individual cells of varying types: independent realizations of the underlying dynamical process governing gene expression. We can represent the state of this entire *eco-community* in terms of the population size for the cells in each type $n(\mathbf{x})$. Different values of $\mathbf{x}$ may represent heterogeneity within the observed population, both in terms of the variability of cells of the same given "type" (i.e., spread within a basin of attraction), as well as the existence of multiple distinct types (i.e., multiple attractors) [66, 149]. The relative depths of the basins of these attractors also gives information about the relative stability of states over different timescales. For example, a "rugged" landscape may have multiple shallow basins along the "walls" of a much deeper basin. Compared to the metastable state at the bottom of the deeper well, the states corresponding to these shallow basins are only transiently stable (i.e., high probability of escape on a long time scale).

Going forward, we make the simplifying assumption that the cell "types" are discrete, biologically discernible objects, which we will call *cell states*. The quantity $n(\mathbf{x})$ is thus a discrete number distribution over a countable number of cell states $\mathbf{x}^{(i)}$: $n(\mathbf{x}^{(i)}) = n_i$. This multi-modal distribution $\{n_i\}$ describes a heterogeneous cell population in terms of its population composition of the subpopulations of its cell states, with cells distributed among different states $\mathbf{x}$ each representing a discernible subpopulation of a size defined by the cell number count $n(\mathbf{x})$. With these premises, the cell population dynamics are given as a balance of cell number in a population over a given time interval:

$$\text{change of \# of cells in type } i = \sum_{j \neq i} \Big(\text{influx type } j \text{ to type } i - \text{outflux type } i \text{ to type } j \Big) + \text{cell type } i \text{ growth}\,. \tag{2.36}$$

The average rates of influx and outflux are given by the barrier heights of the quasi-potential, as in the expression in (1.4).

In traditional non-equilibrium physics, a system in which a non-zero influx balances a non-zero outflux is in a "steady state." However, in biology we can have a growth term due to the coarser modeling granularity of "states": the cell number changes due to cell proliferation and death (net growth), thus apparently defying conservation of mass. In mathematical terms, the infinitesimal generator matrix of the dynamics is no longer Markovian, and its largest eigenvalue is not necessarily zero [148]. Such a system will not reach a stationary population, but only a stationary *proportion* that corresponds to the largest mode of growth (see, e.g., §3.2.1 and §B.2) [59, 66, 148]. This well-known phenomenon, popularized by Gupta and Lander as 'phenotypic equilibrium,' has also been quantified by Zhou, Qian and Huang [53, 148]. For discrete subpopulations, the mathematical existence of such an equilibrium in a generalized branching process was proved by Wang *et al.* and studied experimentally by Pisco *et al.* [70, 98].

2.5.1 Linear population balance model for cell cultures

The continuous-time, linear version of the population balance equation in (2.36) is as follows. Let $\mathbf{n} = (n_1, n_2, \ldots, n_k)$, be the sizes of the k possible subpopulations (cell types). Then we have:

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = \lambda_i n_i + \sum_{j=1}^{k} \nu_{ji} n_j = \sum_{j=1}^{k} a_{ij} n_j \tag{2.37}$$

where the constants ν_{ij} represent the transition rate from subpopulation j to i, with $\nu_{ii} = -\sum_{j \neq i} \nu_{ij}$, and λ_i is the growth rate of type i cells. We define the matrix A as $\Lambda + M$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$, and $M_{ij} = \nu_{ji}$. This dynamical system can be though of as the average dynamics of a continuous time Markov chain (birth/death/state-transition process) with infinitesimal generator matrix A (see §5.3).

The fundamental solution to (2.37) can be expressed as $\mathbf{n}(t) = R(t)\,\mathbf{n}(0)$; the vector $\mathbf{n}(0)$ represents the initial number of cells of each subtype, and the matrix $R(t)$ is given by the matrix exponential $R(t) = \exp\big(tA\big)$. In §3.4, we give an example of the linear population balance model in the context of cancer treatment, which is used to formalize the eco-evolutionary principles of treatment failure due to treatment-induced drug resistance. In §5.1, we discuss how the elements of this rate matrix can be read off directly from single-cell gene expression data derived from experiments in which cell *lineages* are tracked, thereby giving a naïve estimate of the transition rates ν_{ij}. Finally, in §5.3 we lay out how these estimates can be improved upon via a maximum likelihood method adapted from recently published work by Gunarsson *et al.* [52]. Obtaining these estimates is one way in which we attempt to connect the dynamical systems theory of "cell state" to the experimentally observed behavior of cell populations.

Chapter 3

EVOLUTIONARY DYNAMICS OF CANCER PROGRESSION AND TREATMENT

Up to this point, we have given an exhaustive treatment of a quasi-potential landscape for an arbitrary dynamical system, but have not placed it explicitly in the context of understanding cell development and phenotypic heterogeneity. In this section, we aim to do exactly that. In particular, we will discuss the place of landscape theory in the broader realm of cancer modeling.

3.1 Phenotypic heterogeneity and "cancer attractors"

In the past several decades, attempts at improving treatment of cancer have focused on understanding the mechanisms of oncogenesis and metastasis [55]. Research in this vein has illuminated the molecular, genetic, and phenotypic complexity of cancer as a disease, creating the need for a *systems*-level approach to make sense of this complexity [55]. Towards the goal of finding unifying principles that drive cancer progression, biologists D. Hanahan and R. Weinberg proposed six "hallmark traits" that normal cells acquire in the process of becoming cancer cells (Figure 3.1) [55].

In addition to proposing these universal traits, Hanahan and Weinberg advocated for viewing tumors as heterogeneous tissues consisting of multiple cancerous phenotypes, which are in communication with "neutral" ancillary cells of the tumor micro-environment (Figure 3.1) [55, 56]. From the landscape perspective of cell types as attractor states, this intra-tumor heterogeneity is the result of cancer being one of many possible cell fates of a non-cancer progenitor cell [64, 71]. This statement is exactly the theory of a "cancer attractor" outlined in §1.4. Although the theory is couched mathematics, there is significant experimental

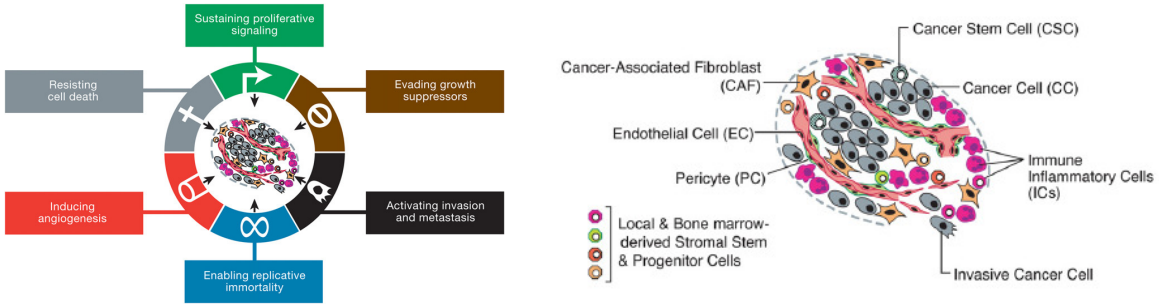

Figure 3.1: *Hanahan and Weinberg's hallmark traits of cancer cells.* **Left.** In their 2000 and 2011 review papers, biologists D. Hanahan and R. Weinberg propose six mechanism-independent characteristics that almost all cancer cells acquire in the process of oncogenesis. These traits allow the cancer cells to sustain proliferation while avoiding targeted cell death. **Right.** Hanahan and Weinberg emphasized the role of phenotypic heterogeneity and cell-cell interaction within tumors in oncogenesis. [55, 56]

evidence that validates the existence of a cancer attractor [64, 65, 79].

3.2 Mathematical models for cancer: mechanistic vs. descriptive

In tandem to experimental progress, recent years have seen increased interest in mathematical models that describe the dynamic heterogeneity of tumor and its micro-environment [3]. Biologist Vito Quaranta separates these models into two broad classes: "mechanistic" and "descriptive." [3]. The former focuses on specific phenomena in cancer progression with the aim of better understanding the underlying biology on a detailed level [3]. The latter concerns itself with the "gross characteristics" (i.e, macroscopic behavior such as size or number of cells) of a tumor, and typically seeks to understand cancer cells from a population dynamic perspective [3].

An example of a mechanistic model is that presented by Anderson and Quaranta (2006) [4]. This model is a multiscale model that encompasses cancer cell motility, proliferation, and evolution with the goal of predicting cancer metastasis and invasion. Mathematically, it has both discrete and continuous elements whose spatiotemporal dynamics obey a set of both probabilistic and deterministic equations, depending on the scale within the model.

An example of a descriptive model is the ODE model proposed by Zhou *et al.* (2014) for a biological population with m intra-converting subpopulations [148]. The goal of this model is to broadly characterize the dynamics of such a population via traditional stability analysis for ODEs, but Zhou *et al.* also give examples of this general model applied to specific cancer cell populations. We use the general ODE set-up from this paper as the basis for our model of drug-resistance in §3.4.

3.2.1 Evolutionary game theory

An important subclass of descriptive models are those that draw on *evolutionary game theory*: the application of game theory to evolving biological populations. The theory relates the evolutionary concept of fitness to the payoff that an individual of a particular subpopulation (e.g., genotype, phenotype, species) receives in a two-player game with another individual [59]. Specifically, it does so through the *replicator equation*, an ODE for the fraction constituted by a given subpopulation within the total population.

More formally, suppose we have a biological population N distinct subpopulations. Let x_i denote the fraction that subpopulation i constitutes in the total population, i.e., the number of individuals of type i divided by the total number of individuals. As $x_i \geqslant 0$ for all i, and $x_1 + \cdots + x_N = 1$, the state variable $\mathbf{x} = (x_1, \ldots, x_N)$ lies in the N-dimensional simplex S^N [59]. Furthermore, the dynamics governing x_i are given by the *replicator equation*:

$$\frac{dx_i}{dt} = x_i\big[f_i(\mathbf{x}) - \langle f\rangle(\mathbf{x})\big], \quad \langle f\rangle(\mathbf{x}) = \sum_{i=1}^{N} x_i f_i(\mathbf{x}), \tag{3.1}$$

where $f_i(\mathbf{x})$ is the *fitness* of an subpopulation i [59]. We call the function $\langle f\rangle(\mathbf{x})$ the *average fitness* of the population [59]. The replicator equation tells us that subpopulation i increases when its fitness is greater than average ($f_i(\mathbf{x}) > \langle f\rangle(\mathbf{x})$), and decreases when its fitness is less than average ($f_i(\mathbf{x}) < \langle f\rangle(\mathbf{x})$).

Evolutionary game theoretic models are particularly advantageous in the study of cancer population biology for a few reasons. By considering the dynamics of interacting *sub*populations,

the replicator equation immediately lends itself to the study of *heterogeneous* populations. In the context of cancer cell populations, the replicator equation has been used to study the effect that competition and cooperation between distinct types of cancer cells has on the efficacy of drug treatment [45, 94, 141, 144]. The replicator equation can admit fixed points in x_i even when the total population tends to infinity, as it often does in models of cancer progression and relapse. These fixed points are easily interpreted in the language of evolutionary biology, as they either occur when a subpopulation goes extinct ($x_i = 0$) or when a subpopulation's fitness is equal to the population average ($f_i(\mathbf{x}) = \langle f \rangle(\mathbf{x})$). In this sense, evolutionary game theory allows for the study and characterization of stable patterns of phenotypic heterogeneity within tumors.

3.3 Drug resistance poses a major challenge to improving cancer treatment

This section, the remaining sections in this Chapter, and §B have previously appeared as the manuscript "A model for the intrinsic limit of cancer therapy: duality of treatment-induced cell death and treatment-induced stemness" (co-authored with Yue Wang, Joseph Zhou, Hong Qian, and Sui Huang), published 2022 in the journal PLOS Computational Biology [6]. All code used to generate the figures in the following sections, as well as those in §B, can be found in the following GitHub repository: https://github.com/eeangelini/cancer-model-plots.

The single major cause of treatment failure in cancer therapy is the emergence of a treatment-resistant tumor that drives recurrence. Other than in the case of some early-stage tumors, it is tacitly accepted that relapse is inevitable during the course of drug treatment. This assumption has translated into the unquestioned practice of quantifying the efficacy of all treatments by how long one extends the time period between diagnosis and either progression, in the form of a clinical relapse, or death [84]. The former metric defines a progression-free survival (PFS) time, which quantifies not the prevention of progression, but a delay, as evident in the proportional hazard model [27]. Treatment success is therefore measured as a "shift to the right" of the decaying Kaplan-Meier curve, which represents the

fraction of progression-free surviving patients in the treated cohort compared to that of the control group. The hence derived extension of median PFS time (or loosely equivalent, of the median time to progression, or TTP) has become a de facto measure for success of a therapy [113].

In theoretical and in vitro experimental models of post-treatment regrowth of a tumor cell population, the "recovery time" in which the surviving tumor cells regrow to reach the population threshold present at the onset of treatment is a biological characteristic of the tissue. It can be considered the equivalent of the clinical TTP [141].

Recurrence, or tumor cell regrowth after treatment, is generally thought to be the result of selection in a process of Darwinian somatic evolution: Given sufficient genetic variability in a sufficiently large initial (pretreatment) cell population, it is considered statistically certain (possibly as a result of increased mutation rate in cancer cells) that the population contains cells carrying genomic mutations that confer drug-resistance and stem-like traits [9, 30, 49, 132]. A single cell with such a mutation will survive the treatment and clonally expand, thus driving the tumor regrowth under treatment.

This genetic explanation implicitly acknowledges de facto inevitability of relapse for a certain set of parameters including mutation rate, cell population size and selection pressure [9, 49, 97, 132]. In addition, elimination of drug-sensitive cells by treatment releases drug-resistant cells from spatial and nutritional constraints and facilitates their proliferation, thereby creating an apparent causal link between treatment and recurrence [18, 109, 141, 142, 143].

In recent years, non-genetic cell phenotype plasticity and the resulting cell population heterogeneity has been recognized as a source of the resistant cell phenotype, which could underlie recurrence without implication of genetic mutations [2, 24, 46, 72, 98, 111, 132]. Extensive phenotypic heterogeneity within a population of cells is generated by the variability of an individual cell's biochemical state. Such non-genetic variability emanates, in part, from the ability of the cell's gene regulatory network to produce a diversity of stable gene expression patterns (attractors), resulting in multistability within a single clonal, isogenic

population. Gene expression noise stochastically disperses individual cells in gene expression space, allowing them to occupy a range of these stable cell states. Because such stochasticity of gene expression causes continuous phenotype switching and equilibration of phenotypes, this type of heterogeneity is not subject to the rule of extinction of a phenotype, as is the case with genetic mutation.

The resulting phenotypic diversity of the isogenic cell population is, while also probabilistic, more prevalent than that generated by genetic mutations and it produces distinct, robust, recurrent, and biologically relevant phenotypic sub-states in clonal cell populations [16, 20, 65]. Such sub-states include mesenchymal, persister, or stem-like states, etc., as single-cell RNAseq now amply reveals [46, 87, 88, 120, 123, 125, 126, 129, 140]. Some of these states can confer resistance and are sufficiently robust to be inherited across several cell generations [20, 93]. In this manner, non-genetic probabilistic variation can also drive Darwinian selection of resistant cells, at least for a number of cell generations.

While both genetic and non-genetic variation arise in a probabilistic manner, there is a key difference. Because non-genetic variant attractor states are the result of regulatory mechanisms, they can also be directly induced by environmental signals. Such *instruction* to change gene expression programs in a directed manner by an external input via a deterministic (or strongly biased probabilistic) control, as opposed to *selection* of an undirected probabilistic internal change, plays a dominant role in a tumor's acquisition of stem-like drug-resistant cells at short time scales (days) compared to the clonal expansion of rare mutant clones [97]. Such regulated change of cell state may be part of a robust, evolved cellular defense mechanism against cellular toxins [63].

A growing body of evidence suggests that emergence of stem-like and therapy-resistant cells along with the associated changes in gene expression are induced (as opposed to selected) by the cytotoxic stress of treatment [121, 125, 137]. In other words, there is a causal biological mechanism linking treatment to stress to the stem-like phenotype. The recurring appearance of stem-cell-like gene expression programs, or "stemness," the speed of response and involvement of canonical signaling pathways that confer multidrug resistance (such as

Wnt signaling-mediated upregulation of the ABC membrane pumps) collectively support the idea of stress-induced activation of preexisting xenobiotic resistance programs in cells by treatment [14, 25, 68].

Therefore, any cytocidal treatment may be a double-edge sword: while a one fraction of cells in the heterogeneous population is killed, another fraction of cells is induced by treatment stress to enter a stem-like or more resistant persister state – planting the seed for recurrence [62, 132]. Drug-induced resistance thus poses an intrinsic limit to curability of tumors under any treatment that involves cell stress, as is the case with chemotherapy, target-selected therapy or radiation. The role of somatic Darwinian evolution in recurrence relative to that of non-genetic cell state transitions has been analyzed using mathematical models in order to minimize selection pressure during treatment [4, 9, 50, 130, 141]. However, the intrinsic limit that induced resistance places on therapy has only recently been systematically evaluated [50].

3.4 A dynamical model of drug-induced treatment resistance in cancer

Here we analyze a most elementary, generic, mathematical model for the conflicting processes that are inherent to cancer therapy: treatment-induced cell death and treatment-induced transition from a drug-sensitive phenotype to a drug-resistant (stem-like) one. Under this formulation, recurrence is "wired-into" the population dynamics and, depending on quantitative details, can become manifest. Over a relevant parameter space of an ordinary differential equation (ODE) model, we analytically evaluate the activity profiles of a drug in inducing cell death vs. transition to the resistant state. We quantify how these features of treatment relate to the intrinsic inevitability of recurrence, measured as TTP. We thus provide a formal survey of the consequence of the core process of non-genetic induction of resistance by treatment, irrespective of the ensuing selection and micro-environmental influences.

We find that depending on the relative susceptibilities for cell death versus induction of the resistant state there can be a non-monotonic dependence of TTP on drug dose, such that beyond a narrow optimal dose, the induction of resistance dominates and increasing treat-

ment intensity will drastically shorten TTP. Thus, knowledge of the measurable propensities of cells to die or activate resistance mechanisms as a function of dose is critical information for optimizing therapy. Our focus on treatment-induced non-genetic tumor cell state change complements the evolutionary framework, fills a conceptual gap to help explain why it is so hard to cure advanced cancer and can be used for modeling scheduling to avoid treatment-associated progression as sought by adaptive therapy [143].

We consider an ODE model that describes the population dynamics of cancer cells that interconvert between two distinct cell states: drug-sensitive and drug-resistant. Let $\mathbf{x}(t) = \big(x_1(t),\, x_2(t)\big)^T$ denote the population state vector, where $x_1(t)$ is the number of sensitive cells, and $x_2(t)$ is the number of resistant cells. Following the formulation given by Zhou et al., the sizes of these two populations are governed by the following linear ODE (Figure 3.2) [148]:

$$\frac{d\mathbf{x}}{dt} = A\,\mathbf{x}, \quad A = \begin{pmatrix} b_S - d_S - k_{SR} & k_{RS} \\ k_{SR} & b_R - d_R - k_{RS} \end{pmatrix} \tag{3.2}$$

We analyze this model of ongoing treatment with the assumption that treatment acts by raising the per capita death rate of cancer cells. Herein, the parameter b_S (b_R) denotes the fixed division rate of sensitive (resistant) cells, whereas d_S (d_R) denotes the total death of sensitive (resistant) cells undergoing drug treatment. The parameter k_{SR} denotes the transition rate at which sensitive cells become resistant, and k_{RS} is the transition rate at which resistant cells become sensitive. All of these parameters are strictly positive and depend in a particular way from the drug dose as discussed below.

We use TTP – the time it takes for the tumor to surpass its pretreatment population size – as a quantitative measure of recurrence under drug treatment. In this model, we denote TTP by t_P, which is defined as

$$t_P = \inf\{t > 0 \,|\, N(t) > N(0)\} \tag{3.3}$$

where $N(t) = x_1(t) + x_2(t)$ is the total number of cells in the tumor (Figure 3.2). The goal

of this model is to understand how t_P depends on the model parameters.

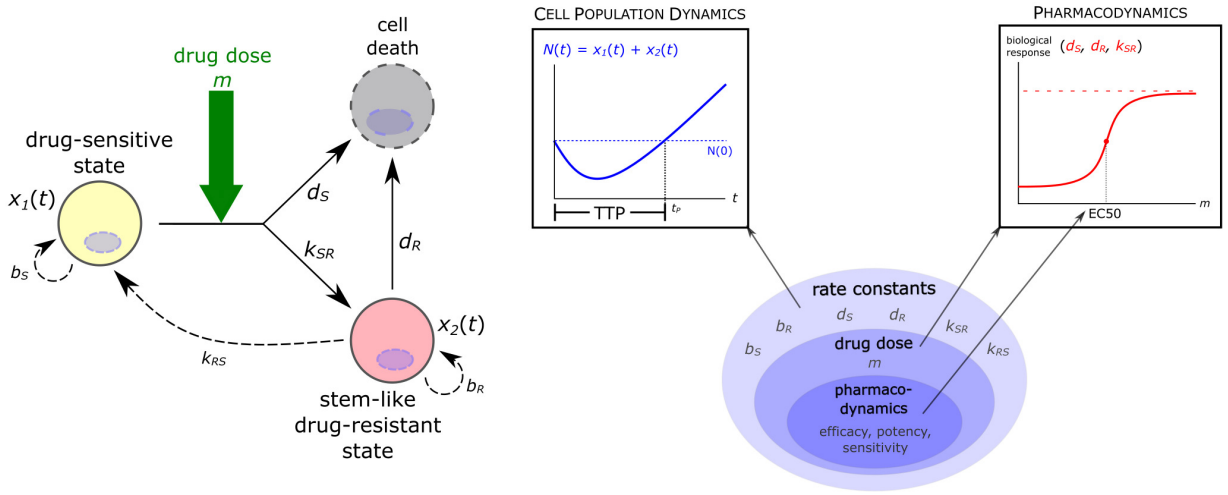

Figure 3.2: *Levels of parameterization in the dynamical model of tumor growth.* At the "highest" (i.e., most general) level, there are the rate constants that govern the growth and state transition dynamics of the cancer cell population. One level down, we introduce drug treatment into the model by assuming that the rate constants d_S, d_R, and k_{SR} are logistic functions of drug dose m. The parameters that determine the shape of these dose-response curves – the pharmacodynamic constants, e.g., EC50 – form the "lowest" (i.e., most specific) level of the model parameters.

3.4.1 Stability analysis

Aside from the degenerate case in which one eigenvalue of A is zero, the only fixed point of the ODE in (3.2) is the point $x = 0$: extinction of all cancer cells. The eigenvalues $\lambda_{1,2}$ of A determine the local stability of this fixed point, whereas the eigenvectors $\mathbf{v}^{(1,2)}$ of A give the primary directions of growth and/or decay in state space. For this model, we can derive general conditions on the growth and transition rates under which the origin is an unstable node, a stable node, or a saddle point (Table B.1). Under appropriate initial conditions, these three cases correspond to unchecked tumor growth, tumor extinction, and tumor regression followed by regrowth. The last scenario provides the simplest mathematical conception for the relapse phenomenon.

The saddle point nature of the extinction state suggests that the dynamics are ultimately difficult to control and contain. It also points to a marked "turning point" for each trajectory: if the initial tumor population is "close enough" to the stable manifold, the trajectory will first move towards the origin before being repelled away along the unstable manifold, indicating the difficulty to eradicate the tumor (Figures B.1-B.4). We can think of this behavior as temporary control of tumor size before the inevitable progression. Equation (B.3) in §B.1 tells us that it is possible to control tumor regrowth wherever the relative fitness of the sensitive phenotype is sufficiently large.

The extinction state is a saddle point whenever the ratio $(b_S - d_S)/(b_R - d_R)$ is greater than the ratio $\frac{(b_S - d_S - k_{SR})}{k_{RS}}$ (Table B.1). That is, temporary control of the tumor size is possible when the relative fitness of the sensitive phenotype is sufficiently large. The constant $b_S - d_S - k_{SR}$ is the net flux of the drug-sensitive population per unit density of drug-sensitive cells, and the constant k_{RS} is the flux into the drug-sensitive population per unit density of drug-resistant cells. The latter parameter, sometimes referred to as the "backflow rate," is useful in characterizing how the pool of drug-resistant cells allows the drug-sensitive population to avoid extinction during chemotherapy [148].

3.4.2 Time to progression

The behavior of TTP as a function of drug dose is closely related to the saddle point dynamics of the system. For one, to have $t_P > 0$, the total population must decrease initially before recovering to its initial value, which is typically only observed when the origin is a saddle point. Moreover, under the saddle point dynamics, $t = t_P$ is the unique non-zero time point for which $N(t) = N(0)$.

Even with a closed expression for $N(t)$, solving the expression $N(t_P) = N(0)$ for t_P is not generally possible as it involves the sum of distinct exponential terms. Instead, we can approximate t_P by decomposing the saddle point dynamics into the distinct stages of population decrease (remission) and increase (regrowth). By definition of TTP, the total population $N(t)$ is undergoing regrowth at time $t = t_P$, i.e., $N'(t_P) > 0$. The theory of linear

dynamical systems tells us that the exact solution $N(t)$ is given by a linear combination of the exponential terms $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$, where $\lambda_{1,2}$ are the eigenvalues of A (as per equation (B.5) in §B.2). In the case of a saddle point, where $\lambda_1 > 0$ and $\lambda_2 < 0$, the tumor population regrowth is necessarily driven by the exponential term corresponding to the positive eigenvalue λ_1. Therefore, at time t_P, we may assume that the total population is well approximated by this exponential term: $N(t_P) \sim e^{\lambda_1 t_P}$. Under this assumption, we can approximate TTP as follows (see §B.2):

$$t_P \approx t_P^* := \frac{1}{\lambda_1} \ln \left[\frac{N(0)}{c_1 \left(v_1^{(1)} + v_2^{(1)} \right)} \right] \tag{3.4}$$

3.5 Pharmacodynamic model of continuous therapy

To consider drug treatment, we assume that drug dose is constant throughout the course of therapy (i.e., continuous therapy) and that the rate constants d_S, d_R, and k_{SR} depend on the amount of drug m present in the system, i.e., that chemotherapy reduces tumor burden by increasing the death rate of cancer cells. At the same time, it increases the rate at which sensitive cells become resistant, the basis for the "double-edged sword" effect of chemotherapy. In mathematical terms, we introduce a secondary parameter m that denotes the drug dose, and we assume that the primary parameters d_S, d_R, and k_{SR} that capture tumor cell population dynamics are increasing functions of m (Figure 3.2). Scaling m as percentage of the maximum tolerated dose (MTD), i.e., $0 \leqslant m \leqslant 100$, we next discuss the pharmacodynamics of cellular drug response, i.e., functional forms for the dependence of the three primary parameters from m.

Using the commonly observed sigmoid shape of biological response curves, which reflect the cumulative probabilistic response of individual cells in a heterogeneous population, we use logistic functions to describe the rate constants $d_S(m)$, $d_R(m)$, and $k_{SR}(m)$ (Figure 3.2):

$$\begin{cases} d_S(m) = \delta_S + \frac{E_S}{1+\exp(-r_S(m-P_S))} \\ d_R(m) = \delta_R + \frac{E_R}{1+\exp(-r_R(m-P_R))} \\ k_{SR}(m) = \kappa + \frac{E_{SR}}{1+\exp(-r_{SR}(m-P_{SR}))} \end{cases} \tag{3.5}$$

We assume that k_{RS}, the rate constant for the re-sensitization of resistant cells, does not change with drug dose.

In the case of $d_S(m)$, each of the four parameters δ_S, E_S, r_S and P_S determines the shape of the logistic curve and describes a behavior affected by the drug. For this reason, we refer to these parameters as the *pharmacodynamic parameters* associated with a given rate constant for cell response. In particular, the parameters E, P, and r respectively determine the maximum response (*efficacy*), EC50 (which is inversely related to *potency*), and saturation rates for each of the above responses of the drug. For example, a drug with high efficacy and high potency to kill sensitive cells is characterized by a high value for E_S and a low value for P_S.

Taken together, our model incorporates three levels of parameterization: first, the parameters of the general linear population dynamics model are the rate constants b_S, b_R, etc. (Figure 3.2). Second, some of these rate constants (e.g., k_{SR}) depend on drug dose, which is represented by the parameter m (Figure 3.2). And third, the way each of these rate constants depend on drug dose is determined by their respective pharmacodynamic constants, as given in (3.5) (Figure 3.2).

In drug development implicit pharmacodynamic parameters are empirically tuned to optimize treatment outcome, which, in the case of our model, is t_P. Teasing apart how the drug dose m affects TTP as a function of the pharmacodynamic parameters is not a

straightforward problem because we do not have an explicit expression for t_P. An exhaustive search of the entire parameter space is impractical. Because drug-induced transition to the resistant state is a hitherto unaccounted-for factor affecting recurrence, we start with a cursory analysis of how t_P changes when we vary the dose-response relationship of the sensitive-to-resistant transition rate $k_{SR}(m)$.

3.6 Qualitative sensitivity analysis of pharmacodynamic parameters

Representative samples of the parameter space demonstrate the possible qualitatively distinct treatment outcomes – t_P as a function of drug dose – that depend on the pharmacodynamics (i.e., rate of behavior as function of drug dose m) of drug-induced resistance relative to killing (Figure 3.3). We define "drug resistance" by taking $E_R \ll E_S$: the drug's cell-killing efficacy is lower for the drug-resistant phenotype than it is for the drug-sensitive phenotype. Although drug-resistance can be manifest in dampening the killing effect either by decreasing the drug potency (or equivalently, increasing the EC50) or efficacy to kill cells (or a combination of both), we do not vary this aspect of resistance. Instead, we model resistance as a lowering the resistant cell death rate at MTD, which reflects a more profound effect on the cell state. To incorporate the observed fitness cost of resistance in the absence of drug, we further assume that $b_S > b_R$ and $\delta_S = \delta_R$ [40, 109, 124, 141].

Thus, we anchor the pharmacodynamics for killing of sensitive and resistant cells $d_S(m)$ and $d_R(m)$ and vary the parameters that determine drug-induced transition to the resistant state with respect to either EC50 (P_{SR}) or efficacy (E_{SR}) for the transition rate $k_{SR}(m)$ (Table B.2, Figure 3.3). We consider these two parameters specifically because of the different effects that they exert on the dose-response curve of a given drug. The efficacy of a drug, or the drug response at MTD, is an intrinsic property of the drug and cannot be compensated by altering the drug dose. On the other hand, the potency of a drug, which corresponds to the inverse of the EC50, is a property that can be compensated by altering drug dose: a low-potency drug can achieve the same response as a high-potency drug, but at a higher drug dose.

The unknown then is how the relation between rate of induction of the resistant phenotype, $k_{SR}(m)$, and the "kill curves," $d_S(m)$ and $d_S(m)$, shape t_P as a function of drug dose m. The coarse-grained, but comprehensive, sensitivity analysis of our model model is achieved by altering the values of P_{SR} and E_{SR}, relative to the two fixed kill curves pharmacodynamics curves (Figure 3.3). The following four canonical cases represent qualitatively distinct, plausible pharmacodynamical relationships between killing and induction of resistance due to treatment.

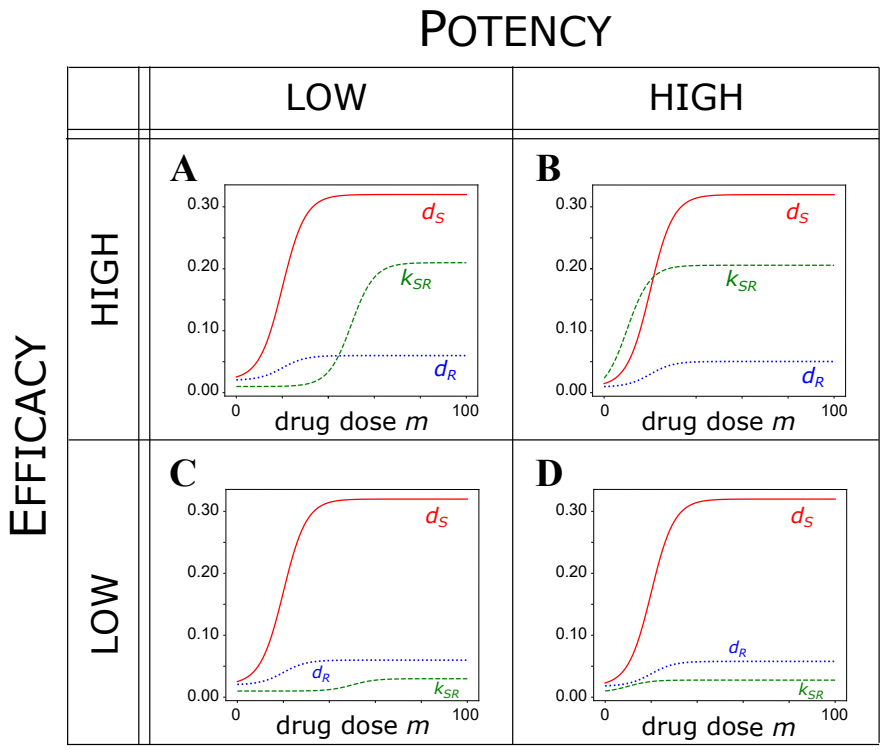

Figure 3.3: *A coarse-grained sensitivity analysis of the effect of drug-induced resistance on treatment outcome.* The parameters given in Table B.2 are held fixed, while the EC50 and efficacy of resistance induction are varied in the above four cases. Efficacy increases as the parameter E_{SR} increases, whereas potency decreases as the EC50 value (P_{SR}) increases. **A.** Case A: high efficacy, low potency ($E_{SR} = 0.2, P_{SR} = 50$). **B.** Case B: high efficacy, high potency ($E_{SR} = 0.2, P_{SR} = 10$). **C.** Case C: low efficacy, low potency ($E_{SR} = 0.02, P_{SR} = 50$). **D.** Case D: low efficacy, high potency ($E_{SR} = 0.02, P_{SR} = 10$).

3.6.1 Case A: high efficacy, low potency of resistance induction

We first consider a drug with a high efficacy and low potency (i.e., high E_{SR} and P_{SR}) for inducing resistance (Figure 3.3A). For such a drug, t_P is not monotone increasing in drug dose. Instead, it first increases before decreasing to a plateau (Figure 3.4A). Thus, increasing drug dose past a certain point significantly worsens the treatment outcome.

We observe that t_P increases with drug dose when an increase in dose corresponds to a significant decrease in the total growth rate relative to the total switching rate (Figure 3.4A). On the other hand, t_P decreases with drug dose when an increase in drug dose corresponds to a significant increase in the total switching rate relative to the total growth rate (Figure 3.4A). The fact that the transition dynamics dominate the growth dynamics at high drug doses indicates that under cytotoxic stress, tumor recurrence is not driven by cell growth (Figure 3.4). Instead, it is driven by the ability of the tumor cells to evade extinction by transitioning to a stem-like, drug-resistant state.

Returning to our analysis of the saddle point dynamics, the positive eigenvalue λ_1 is also non-monotonic in drug dose, first decreasing to a minimum before increasing to a plateau (Figure 3.4). This inverse relationship between λ_1 and t_P agrees with the above observation that λ_1 roughly corresponds to the rate of tumor growth during recurrence.

3.6.2 Case B: high efficacy, high potency of resistance induction

We now consider a drug with the same efficacy in inducing the stem-like state as in Case A but with higher potency (Figure 3.3B). Unlike Case A, t_P is now monotone increasing in drug dose (Figure 3.4B). Therefore, if the drug dose is sufficiently high, the treatment outcome in Case B is not sensitive to fluctuations about a dosage. This robustness is favorable to the sensitivity observed under Case A, in which a small fluctuation in drug dose can significantly worsen the t_P. On the other hand, the maximum t_P in Case B is low compared to that in Case A (Figure 3.4A-B). A treatment that is robust to variation in drug dose is not necessarily a good one if progression is only, at most, slightly delayed.

As before, the behavior of t_P is well summarized by the positive eigenvalue λ_1 and the difference between the total growth and switching rates (Figure 3.4B). In particular, the latter shows how increased drug potency to induce stemness affects t_P. As EC50 for k_{SR} is closer to that for d_S and d_R than it is in Case A, the total growth and state-switching rates saturate at roughly the same drug dose (Figure 3.4B). Therefore, the difference between the two rates only increases over a small range of doses, in which in t_P increases monotonically (Figure 3.4B). Cases A and B reveal that the relative potency of a given drug to induce resistance versus kill cells determines whether t_P is monotone increasing in drug dose or not.

3.6.3 Case C: low efficacy, low potency of resistance induction

For a drug with the same low potency as in Case A, but with lower efficacy, t_P is again, as in Case A, a non-monotonic function of drug dose (Figures 3.3C, 3.4C). However, the difference between the maximum possible t_P and t_P at MTD is much smaller for the drug than in Case A, since the maximum difference between the total growth and switching rates is smaller than in Case A (Figure 3.4A, 3.4C). In terms of treatment outcomes, we can think of this drug as an improvement from Cases A and B: t_P is overall less sensitive to variation in drug dose than in Case A, and the maximum possible t_P is greater than that in both Cases A and B.

The overall increase in t_P is also reflected in saddle point dynamics of the system. In Case C, the magnitude of the negative eigenvalue λ_2 is overall lower than in Case A (Figure 3.4C). This decrease in magnitude agrees with the increase in t_P, as λ_2 roughly corresponds to the rate of tumor remission in the early stages of treatment. That is, the remission stage is prolonged under treatment by a low-efficacy drug, as compared to a high-efficacy drug (Figures B.1-B.4).

3.6.4 Case D: low efficacy, high potency of resistance induction

Finally, we consider a drug that is a "combination" of Cases B and C; that is, the drug has low efficacy but high potency in inducing resistance to cell killing (Figure 3.3D). Under this

drug, t_P is a similar "combination" of Cases B and C: it is monotone increasing in drug dose, and its value at MTD is the same as in Case C (Figure 3.4D). Therefore, a drug with low efficacy and high potency for drug-induced resistance relative to killing produces a treatment scheme that effectively delays tumor progression for a wide range of drug doses. Compared to case C, the higher potency in inducing stemness at low dose abrogates the optimal dose window, i.e., the counterintuitive peak in t_P at lower dose.

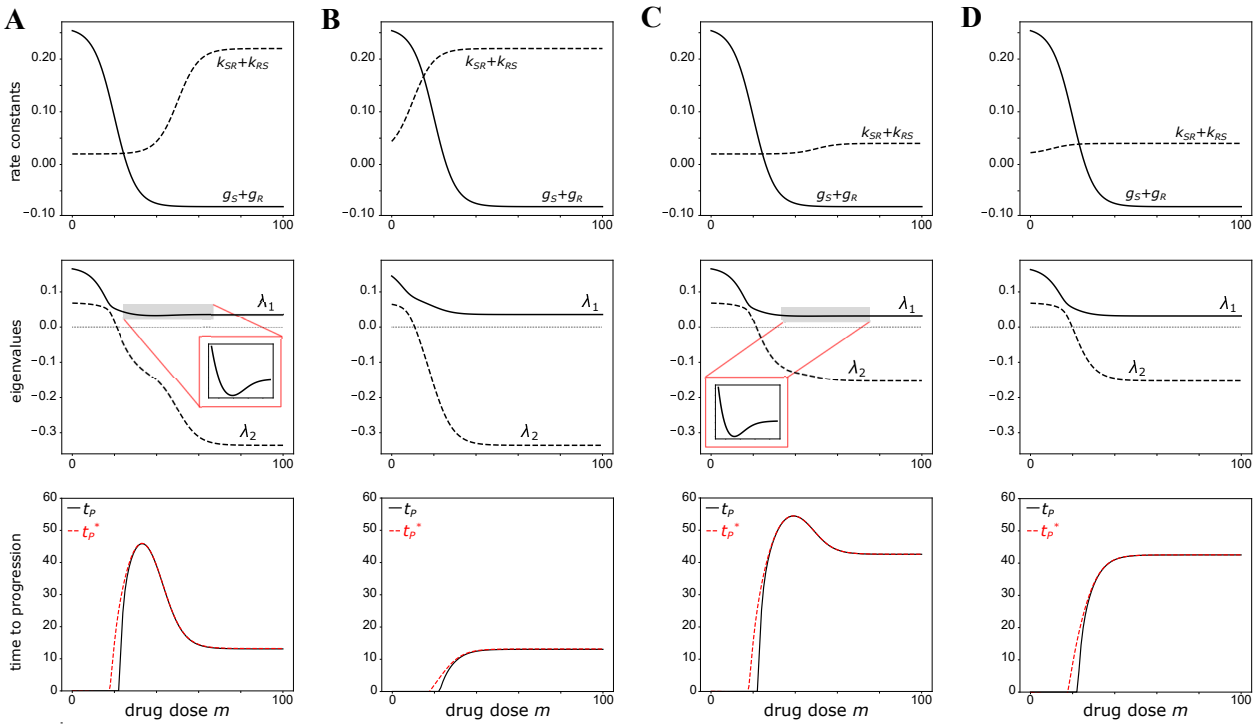

Figure 3.4: *Summary of model behavior for Cases* ***A***, ***B***, ***C***, *and* ***D*** *as a function of drug dose* m. The net growth rates of sensitive and resistant cells are defined as $g_S := b_S - d_S$ and $g_R := b_R - d_R$, respectively. *First row:* Total growth and transition rates, $g_S + g_R$ and $k_{SR} + k_{RS}$, respectively. *Second row:* Eigenvalues of the matrix A in (3.2). The inset in cases A and C highlight the non-monotonic behavior of the eigenvalue λ_1. *Third row:* Time to progression t_P plotted alongside its asymptotic approximation t_P^*.

3.6.5 Parameter search

In order to determine how well Cases A, B, C, and D capture the dependence of t_P on drug potency (P_{SR}) and efficacy (E_{SR}) of resistance induction, we perform a broad parameter search. Specifically, we compute t_P over a wide range of P_{SR} and E_{SR} beyond the four representative cases from the previous section (Figure 3.3). Instead of plotting t_P as a function of drug dose m, we plot three quantities of interest as a function of P_{SR} and E_{SR}: the drug dose at which t_P attains its maximum, the maximum value of t_P, and the value of t_P at MTD (Figure 3.5). Doing so, we find that t_P attains its maximum at MTD (i.e., is monotonic in drug dose) when P_{SR} is low regardless of the value of E_{SR} (Figure 3.5A). Therefore, t_P is indeed a non-monotonic function of drug dose only when the potency to induce resistance is low. Furthermore, the maximum value of t_P shows little sensitivity to either E_{SR} or P_{SR}, and instead appears to depend on whether t_P is monotone increasing in drug dose or not (Figure 3.5B). We also find that the value of t_P at MTD decreases as E_{SR} increases independent of the value of P_{SR} (Figure 3.5C).

The results of this finer-grained parameter search indicate that the four cases given in the main text indeed characterize the dependence of t_P on the parameters P_{SR} and E_{SR}: the non-monotonic dependence of t_P on drug dose is governed by P_{SR}, whereas the value of t_P at MTD is governed by E_{SR}. This parameter search, however, is still limited to the two parameters P_{SR} and E_{SR}. Taken together, a key finding is that at low potency (high EC50) for inducing resistance, where higher drug doses m are needed to trigger this cell state transition, increasing the dose past a certain point reduces t_P. A further evaluation of the parameter space, paired with rigorous sensitivity analysis, is required to characterize how t_P depends on the complete set of model parameters.

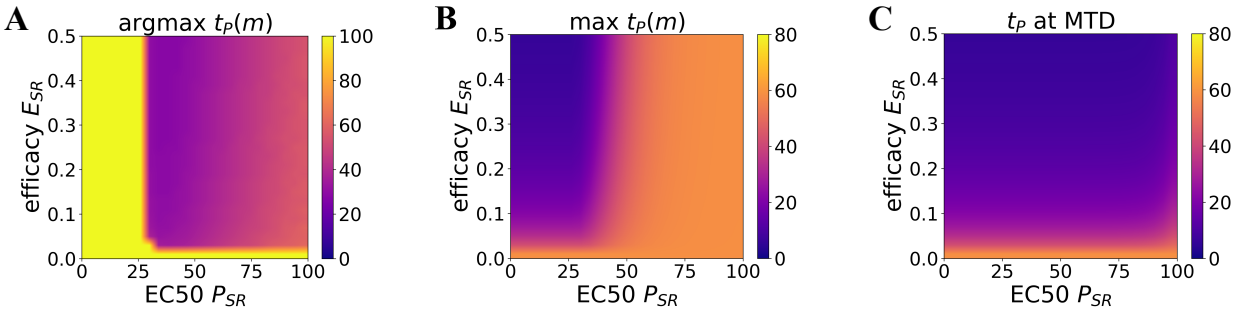

Figure 3.5: *Analysis of time to progression (t_P) over a wide range of values for potency (P_{SR}) and efficacy (E_{SR}) of inducing resistance.* Plotted values are indicated by the colorbar on each plot. All other parameters are fixed at the values given in Table B.2. **A.** Drug dose m at which t_P attains its maximum. **B.** Maximum value of t_P over all drug doses. **C.** Value of t_P at MTD.

3.7 Virtual patient cohort simulations

In statistical analysis of clinical studies, individual patient measures, such as PFS time and TTP, are typically not displayed directly; instead, the data are often presented in the form of Kaplan-Meier curves, which show the fraction of surviving and progression-free patients as a function of time [27, 113]. The process of calibrating mathematical models of cancer dynamics to clinical data typically involves applying regression, or some other parameter-fitting method, to a Kaplan-Meier curve from observed cancer patient cohorts [11, 76]. Doing so requires generating a "virtual patient cohort" from the model, usually by assuming some statistical distribution of a set of the model parameters, and sampling individual "patients" from this distribution [11, 76]. As a step towards grounding our model in clinical data, so as to make meaningful predictions about treatment courses, we generate virtual patient cohorts and present an analysis of the resulting Kaplan-Meier curves.

To generate virtual patient cohorts, we assume that the basal rates of cell birth, death, and state transition, as well as the fraction of resistant cells at tumor detection, vary from patient to patient (Table B.3). For simplicity, we assume that each of these parameters are uniform random variables. One possible way of updating this assumed prior with a more appropriate

posterior distribution based on clinical data would be to use an expectation-maximization approach [76]. We assume that the remaining parameters, tumor size at diagnosis and the pharmacodynamic constants, remain fixed across all patients, as they are intrinsic properties of standard clinical practice and a given drug, respectively (Table B.2).

Keeping with our previous analysis, we generate $n_c = 100$ virtual patient cohorts of $n_p = 10^3$ patients each, and compute the TTP curve $t_P(m)$ for each patient at values of E_{SR} and P_{SR} given by the four previous cases A, B, C and D (Figure 3.3). Once we have $t_P(m)$ for each patient in a given cohort, we can compute the progression-free fraction for a range of possible drug doses, where we treat t_P as a progression event for each patient. We then take the mean progression-free fraction across all cohorts, giving us a set of Kaplan-Meier curves (Figure 3.6). Our aim is to see if the non-monotonic dependence of treatment outcome on drug dose m is manifest at the ensemble level of the patient cohort. Whereas for an individual patient, an improved treatment outcome is indicated by an increased TTP, an improved outcome for an entire cohort is indicated by an upward and/or rightward shift in the Kaplan-Meier curve.

Comparing the resulting Kaplan-Meier curves for cases A-D, we find that the mean behavior of the survival fraction mirrors that of t_P. As before, in cases A and C, where potency to induce resistance is low, the overall treatment outcome is non-monotonic in drug dose: as m increases from 0 to MTD, the Kaplan-Meier curve shift upwards, then downwards (Figure 3.6). Also in line with our TTP analysis is the observation that in cases C and D, where efficacy to induce resistance is low, the overall treatment outcome is better than that for cases A or B: the upwards shift of the Kaplan-Meier curves going from low to high drug doses is much larger in the former cases than in the latter (Figure 3.6). Thus, our conclusions about the qualitative behavior of the model at the level of the individual "patient's" tumor cell population scale up to the level of the statistical ensemble (i.e., the cohort), replicating both the "expected" as well as the "counterintuitive" effects of increasing drug dose. This result is significant as it demonstrates that our analysis is in some sense robust to noise. The next logical step would be to see if our analysis still holds, and is therefore clinically

meaningful, after fitting our parameters to actual patient data.

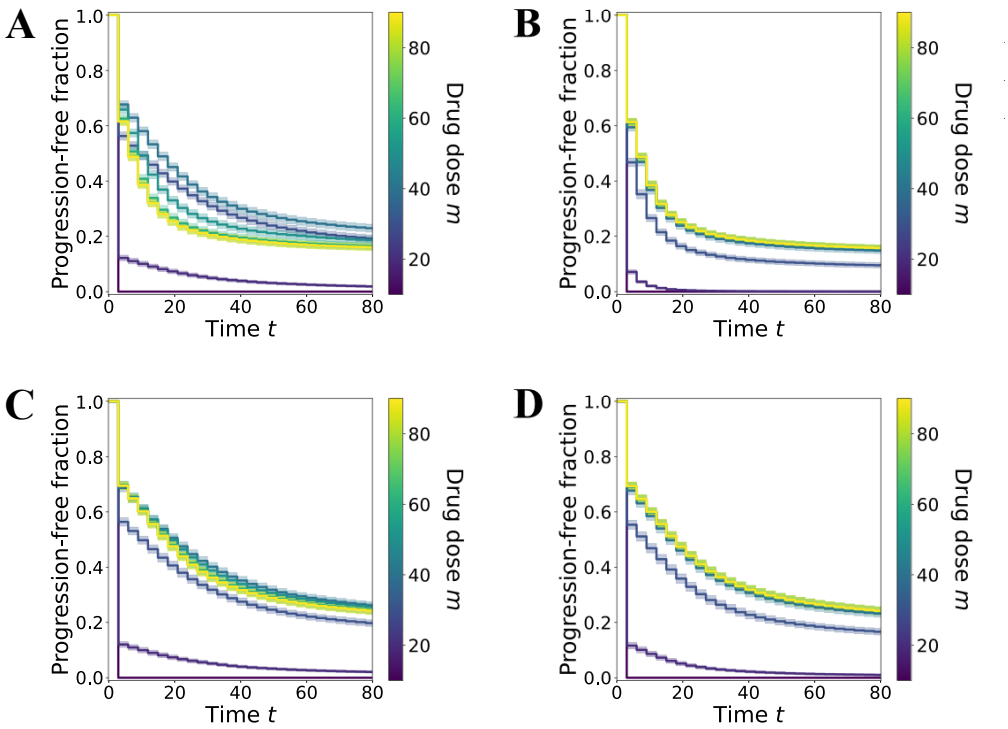

Figure 3.6: *Kaplan-Meier curves averaged over* $n_c = 100$ *virtual cohort simulations of* $n_p = 10^3$ *patients each, plotted for fixed drug doses* m. Error bars denote plus and minus one standard deviation. The x-axis denotes time t, and the y-axis denotes the fraction of patients in a cohort who are progression free by time t (i.e., $t_P > t$). Each curve is colored according to the fixed drug dose m applied to each cohort, indicated by the color bar in each plot. Panels **A**, **B**, **C**, and **D** correspond to varying the parameters P_{SR} and E_{SR} according to the four cases shown in Fig 3.3. Unless specified as being drawn from the uniform distributions given in Table B.3, all other parameters are held fixed at the values given in Table B.2.

3.8 Discussion and conclusions

In advanced tumors, post-treatment recurrence is almost an intrinsic feature in the course of tumor progression. It is increasingly acknowledged that even if an untreated tumor would result in rapid progression and death, recurrence after treatment is causatively or mechanistically linked to the act of treatment. The traditional explanation for recurrence after initial remission invokes selection of preexisting mutant cells in which genetic mutations confer the

resistance phenotype.

More recently, the "competitive release" of the resistant cells, when these cells expand into the niche freed by the killing of sensitive cells by treatment, has been proposed as mechanism of tumor recurrence, adding the non-intuitive twist that "more killing" is not better [18, 50, 109, 141, 142, 143] . Nonlinear models of competition between sensitive and resistant subpopulations, such as that presented by Kozlowska *et al.*, lead to similar saddle-point dynamics of tumor depletion and progression as our model, albeit at a longer time scale that encompasses multiple cell generations [76].

Another interesting fundamental similarity between genetic and non-genetic mechanisms, at least in terms of a possible formal generalization, pertains to an additional layer of non-linearity that we have not considered here: a multi-step process that gives rise to a progressive increase ('gradual evolution') in resilience. Increasing genomic instability with tumor progression not only decreases the threshold for cell death exploited by many therapies, but also accelerates mutational exploration of new phenotypes in an evolutionary process. A well-studied case of self-propelling increase in resistance is the amplification of the DHFR gene repeats that confers resistance to methotrexate. Once amplified, the locus is even more prone to undergo genomic recombinations and to further amplify, including in response to treatment stress that promotes chromosomal breakage [73]. The non-genetic equivalent is that with increasing malignancy ("stemness"), the barrier for cell type transitions is reduced, evident in the increasing phenotypic plasticity of advanced tumors, and hence, the chance of the tumor cell to enter an even more dedifferentiated, resilient stem-like cells in response to treatment stress [73].

The oft observed re-sensitization of recurrent tumors, rapid rate for appearance of resistance markers, and ubiquitous cell phenotype plasticity, however, suggest a role for non-genetic, reversible phenotype switching in tumor recurrence [2, 24, 34, 46, 54, 72, 81, 98, 111]. Most recently, the induction of cells to transition into a stem- or mesenchymal-like resilient state by the cell stress imparted by treatment has received increasing acceptance and has been confirmed by single-cell resolution measurements [16, 20, 46, 65, 87, 88, 120, 123, 125,

126, 129, 140]. Population-wide resistance to treatment induced by the same perturbation intended to kill and eradicate tumor cells thus poses a conflicting situation: a double-edged sword that complicates treatment response.

The mechanisms that allow a cell to be either fated to death or to enter a resistant state following the same perturbation may depend on the initial microstate due to stochastic gene expression [62]. This uncertainty is here modeled by the sigmoidal shape of the pharmacodynamical functions, which represents the cumulative probability of the dispersed propensity of cells to respond in either way to treatment. We used a minimal model to characterize how the "relative strength" (potency and efficacy) of a drug to either kill tumor cells or convert them into resistance cells affects the population dynamics, as measured by the time it takes for the cell population to recover and grow to its pre-treatment size (time to progression t_P). We focused on four qualitatively distinct scenarios (A, B, C, and D) that correspond to all possible combinations of high versus low potency and high versus low efficacy of the treatment to induce the resistant state relative to killing.

Despite the elementary form of the model, interesting behaviors emerge: the four scenarios produced robust, prototypic behaviors for the dependence of t_P on drug dose (Fig 3.5). The two cases (A and C) in which the potency of the drug (dose for half-maximal effect, EC50) to induce the resistant phenotype was substantially lower (high EC values) than the potency to kill cells produced a non-monotonic t_P-to-drug dose relationship. In such a scenario, there is an optimal window of dose for maximal t_P: drug doses higher than the optimal dose will have lower benefits in terms of t_P.

The efficacy of the drug (the amplitude of the dose-response curve) also affected t_P by determining its plateau value at high drug doses. High efficacy of the drug to induce the resistant state (Cases A, B) resulted, as expected, in a low value of t_P at high doses, independent of whether the dose-dependence is monotonic (Case B) or non-monotonic (Case A). On the other hand, a drug with low efficacy to induce treatment resistance (Cases C, D) resulted in a high value of t_P at high doses. Sensitivity of t_P to changes in drug efficacy means that drug dose optimization is paramount in the case where the potency of the drug

to induce resistance is low relative to the potency of cell killing – parameters that could be determined in preclinical studies.

Unfortunately, fine-grained dose-escalation clinical trials (e.g. with at least three doses) are generally not conducted that would expose the non-monotonic effect. To establish the connection between intrinsic biological properties of drugs in triggering state transition to the resistant state and the clinical consequences as observed in drug trials, we performed virtual patient cohort simulations. We found that the qualitative conclusions of the model about the effect of induced resistance on treatment success are robust to variation in the other model parameters and result in corresponding non-monotonic dependence of the progression free survival in the simulated patient drug trial cohorts

Adaptive therapy, in which treatment is stopped upon regression and re-started upon regrowth, or metronomic therapy, which applies a low dose at regular intervals, may be worthwhile treatment schemes in cases A and C because there is an optimal drug dose below MTD in these cases. The current rationale behind dose-minimization treatment strategies is to avoid fixation of mutant resistant clones due to competitive release and selection [4, 18, 109, 142, 143]. However, if resistance is inducible and reversible, the intended "containment" of the tumor (as opposed to the harder-to-achieve "eradication") is even more readily achieved than when guided by the concept of competitive release of mutant resistant clones. Thus, if we consider the new biological rationale of non-genetic, reversible dynamics of treatment-induced resistance, a strategy such as adaptive or metronomic therapy may be further optimized to be more effective. However, in order to make any meaningful predictions about optimal treatment courses we must first ground our model in clinical data, either by using empirical estimates of parameters from the literature, or by using a statistical learning framework to fit parameter distributions from the data [11, 17, 75, 76].

Chapter 4

STOCHASTIC THERMODYNAMICS FOR "BIG DATA"

Following a recently reformulated Gibbs' statistical chemical thermodynamic theory on discrete state space, we present a treatment on statistical measurements of random mechanical motions in continuous space. In particular, we show how the concept of temperature as well as ideal gas/solution law arise from a statistical analysis of a collection of independent and identically distributed complex particles without relying on Newtonian mechanics, nor the very concept of mechanical energy. When sampling from an ergodic system, the data *ad infinitum* limit elucidates how entropy function characterizes randomness among measurements, with a novel energetic representation for the statistics and an internal energy *additivity* emerge. This generalization of Gibbs' theory is applicable to statistical measurements on single living cells and other complex biological organisms, one individual at a time. This Chapter and §C have previously appeared as the manuscript "Statistical analysis of random motion and energetic behavior of counting: Gibbs' theory revisited." (co-authored with Hong Qian), published 2023 in the Journal of Physical Chemistry B [6].

4.1 Gibbsian thermodynamics: from biophysical chemistry to living systems

In this work we attempt to extend the biophysical chemistry paradigm on single proteins, which are essentially statistical measurements combined with Gibbsian thermodynamics and chemical kinetics [116]. In particular, we aim to extend this work from "dead" molecules to living cells by replacing the foundation of current Gibbs' theory, which relies on mechanics and the very concepts of molecular energy and chemical equilibrium, by the mathematical theory of probability.

A recent alternation and generalization of Gibbs' statistical chemical thermodynamics

has identified the following: [104, 105]

(i) Statistical counting as the most fundamental empirical observational activity in a *Gedankenexperiment*, more so than measurements of mean values of real-valued observables. There is missing information in the latter with respect to the former;

(ii) the mathematical theory of probability as a fundamental law for statistical knowledge;[67, 69]

(iii) entropy functions as emergent concept, *à la* P. W. Anderson, which gauges the difference between observed empirical statistics *ad infinitum* in (i), and a prior probability in (ii) ; [5]

(iv) the entropy function for counting frequency provides the foundation of a statistical thermodynamics, *à la* H. B. Callen, that not only dictates the fundamental thermodynamic relation that involves conjugate variables, but also the posterior statistical laws in terms of Gibbs' ensembles [23].

The new theory establishes a foundation for statistical thermodynamics independent of mechanics. It actually provides the very concept of thermodynamic internal energy a new conception. One can see a precise statistical logic that organically bridges the frequentist-Bayesian divide [44]. The theory is applicable to any other statistical knowledge.

In the present work, we turn our attention particularly to mechanical motions with the newfound perspective and method. From a very primitive, empiricism stand point, if chemistry is about "counting the number of occurrence of various states of a molecule", then mechanics in condensed phases is about "counting the number of steps taken, in different directions with different sizes, by a point mass" in the three-dimensional physical space.

The prior probability, as a hypothesis, figures prominently in the new formulation of Gibbs' theory [67, 69]. In traditional statistical mechanics, this prior probability has been taken as the natural (Lebesgue) measure of the state space, known as the postulate of equal

a priori probability. In the light of our current theory, the problem of normalization of the prior probability is the most significant loophole in the earlier theoretical setup. Careful analysis of this problem has been taken in the present work.

All real objects that consist of atoms are dynamic; they can and should be understood as dynamical systems. The mathematical theory of ergodic dynamical systems — be it deterministic, chaotic, or stochastic — is an abstract construct. Nevertheless, this theory provides a theoretical foundation for statistical data science [106]. In reality, every observable object is only metastable, *i.e.*, it exists only on certain time scale and disintegrates over a much longer time scale. For example, even though a protein *in vitro* has a lifetime of a millennium without the help of an enzymatic catalyst, this lifetime is still finite in the mathematical framework. Inside a living cell, however, this time scale is shortened to days or weeks [35, 108]. The statistical study of single protein molecules *in vitro*, therefore, idealizes a protein as an ergodic polypeptide [145].

In the current mathematical theory of molecular cells, there is an idealized ergodic biochemical dynamics of a single cell with a given genome: An "eternal cell" ultimately passes through all its possible biochemical states, with different epigenetic phenotypes being (meta)stable dynamic "attractors" [66, 101]. For a collection of data sampled from different cells with the same genome, it is therefore legitimate to assume that the data are samples from an ergodic dynamical system with a common probability measure despite being snapshots of different cells and not a single time series. For such a system, however, it is not plausible to empirically determine the ergodic, invariant probability measure of the stochastic dynamics of the idealized eternal cell. Therefore, it is important to explicitly acknowledge the assumption of the prior probability, as it will remain only a scientific hypothesis in our understanding of single cells. The theoretical concept of a biochemical, nonequilibrium steady state (NESS) landscape of a cell plays a unique role in this regard [7, 66, 101, 133].

A key assumption of the present work is that statistics of data *ad infinitum* from ergodic samples are *intrinsic properties* of the dynamical system being sampled. Empirical data analysis in a laboratory of multiple samples is best understood as from an identical proba-

bility distribution. In principle, the framework of an ergodic sample *ad infinitum* eliminates all the statistical uncertainties and guarantees reproducibility.

Important insights are derived from the current theory of statistical energetics: Sec. 4.4.1 shows how information entropy maximization under constrains gives rise to a conservation law, in terms of Lagrangian multipliers in calculus as the conjugate variables, that reveals the holographic, total internal energy being decomposed into a "known" and a "missing" parts (Eq. 4.21). In Sec. 4.6, a collection of particles with complex internal state space $\mathcal{X}$ and 3-dimensional spatial movements, but nevertheless statistically independent, is shown to have an ideal-solution-like energetics (Eq. 4.42), even $pV = NT$ with p and T being the respective conjugate variables to the counting statistics of spatial positions and internal states. Under this unorthodox view, a point mass has a simple "internal state space" in terms of its instantaneous velocity.

There are two issues when working with the probability on a continuous state space. At the very basic level, our intuitive understanding of probability stems from normalized counting frequencies, and the very activity of "counting" implies a discrete state space. On the more mathematical level, the probability of any elementary event in a continuous state space is zero; there is no possibility of repeated two measurements with identical real values. To solve this problem, mathematicians invented the concept of probability measure and scientists argued for measurement resolution as a discretization [69]. For continuous mechanical state space, a third issue is the normalizability of probability. This is the motivation of the treatment developed in Sec. 4.2.3, where we restrict the problem to a finite d-dimensional ball, whose radius is arbitrarily large but finite. The present work suggests that the theory of ideal gas and ideal solution implicitly "solved" this problem by constraining all the particles in a finite volume V and their velocities follows a multivariate Gaussian distribution with finite temperature.

4.2 Statistical measurements, entropy, and the Legendre-Fenchel Transform

In this section, we present key mathematical equations and their derivations. The mathematical equations and formulas established in this section are the bases for the results in the sections that follow. One can skip this section and directly go to Sec. 4.3 for continuity of the narrative.

4.2.1 Statistical energetics of counting by the methods of Willard Gibbs

By *energetics*, we mean the theory of entropy and energy. We use a macromolecule with $n+1$ possible conformations to illustrate the mathematical method that first appeared in Gibbs' work [51]. We follow the presentation in refs. [104] and [105] where the logic flow is completely different and the scientific content altered from the current teaching in statistical thermodynamics [13, 57]. Let $\mathcal{S} = \{0, 1, \cdots, n\}$ be the discrete state space of the macromolecule and a very large integer N be the number of statistically independent and identical samples of the single molecule. Under the prior probability $\mathbf{p} = (p_0, \cdots, p_n)$, the relative (negative) entropy function Φ quantifies the infinitesimal probability of observing normalized counting frequency $\boldsymbol{\nu} = (\nu_0, \cdots, \nu_n)$ among the N samples:

$$\Phi(\boldsymbol{\nu}) = \sum_{i \in \mathcal{S}} \nu_i \ln \left(\frac{\nu_i}{p_i} \right). \tag{4.1a}$$

This is a measure of the amount of information, or "surprise," in the counting frequency data $\boldsymbol{\nu}$ w.r.t. the prior probability $\mathbf{p}$ [26]. That is, there is no "surprise" in the observation when $\boldsymbol{\nu} = \mathbf{p}$, and accordingly, $\Phi(\mathbf{p}) = 0$.

The Legendre transform has been described as the "workhorse of chemical thermodynamics" [1]. The modern mathematics of convex analysis has refined this mathematical technique in the form of the Legendre-Fenchel transform (LFT) [110]. One can compute the

LFT of $\Phi(\boldsymbol{\nu})$ as a function of the conjugate variables $\boldsymbol{\mu} = (\mu_0, \cdots, \mu_n)$:

$$\Psi(\boldsymbol{\mu}) = \inf_{\boldsymbol{\nu}} \left\{ \sum_{i\in\mathcal{S}} \nu_i \mu_i + \Phi(\boldsymbol{\nu}) \right\} = \inf_{\boldsymbol{\nu}} \left\{ \sum_{i\in\mathcal{S}} \nu_i \ln\left(\frac{\nu_i}{p_i e^{-\mu_i}}\right) \right\} = -\ln \sum_{i\in\mathcal{S}} p_i e^{-\mu_i}. \tag{4.1b}$$

The optimal $\boldsymbol{\nu}$ is given by $\nu_i^* = p_i e^{-\mu_i}/Z(\boldsymbol{\mu})$ with $Z(\boldsymbol{\mu}) = \sum_{i\in\mathcal{S}} p_i e^{-\mu_i}$. If one interprets the conjugate variables $\boldsymbol{\mu}$ as the statistical energies of each state in units $k_B T$, then Ψ is precisely the free energy in Gibbs' theory [13, 57].

Capturing the relation between the frequency of counting $\boldsymbol{\nu}$ and the chemical-potential-like quantity $\boldsymbol{\mu}$, Fig. 4.1 summaries the overall structure of the new theory proposed in ref. [104]. The four transformations on the left of Fig. 4.1 are all invertible, where the LFTs are

$$\Psi(\boldsymbol{\mu}) = \inf_{\boldsymbol{\nu}} \big\{ \langle \boldsymbol{\mu}, \boldsymbol{\nu} \rangle + \Phi(\boldsymbol{\nu}) \big\}, \;\; \Phi(\boldsymbol{\nu}) = -\inf_{\boldsymbol{\mu}} \big\{ \langle \boldsymbol{\mu}, \boldsymbol{\nu} \rangle - \Psi(\boldsymbol{\mu}) \big\}, \tag{4.2a}$$

$$\Psi_{\boldsymbol{g}}(\mathbf{y}) = \inf_{\mathbf{x}} \big\{ \langle \mathbf{y}, \mathbf{x} \rangle + \Phi_{\boldsymbol{g}}(\mathbf{x}) \big\}, \;\; \Phi_{\boldsymbol{g}}(\mathbf{x}) = -\inf_{\mathbf{y}} \big\{ \langle \mathbf{y}, \mathbf{x} \rangle - \Psi_{\boldsymbol{g}}(\mathbf{y}) \big\}, \tag{4.2b}$$

in which a complete set of observables with holographic information $\mathbf{x} = (1, x_1, \cdots, x_n)$ is represented by a linear transformation with non-singular matrix $\mathbf{G}$, $\mathbf{x} = \boldsymbol{\nu}\mathbf{G}$:

$$\Phi_{\boldsymbol{g}}(\mathbf{x}) = \Phi\big(\mathbf{x}\mathbf{G}^{-1}\big), \;\; \Phi(\boldsymbol{\nu}) = \Phi_{\boldsymbol{g}}(\boldsymbol{\nu}\mathbf{G}), \tag{4.3a}$$

$$\Psi_{\boldsymbol{g}}(\mathbf{y}) = \Psi\big(\mathbf{G}\mathbf{y}\big), \;\; \Psi(\boldsymbol{\mu}) = \Psi_{\boldsymbol{g}}\big(\mathbf{G}^{-1}\boldsymbol{\mu}\big) \; . \tag{4.3b}$$

The two transformations on the right of Fig. 4.1, from the n-vectors $\mathbf{x}$ and $\mathbf{y}$ to the J-vectors $\mathbf{x}'$ and $\mathbf{y}'$, where $J < n$, relate microscopic ("holographic") observables to macroscopic ("coarse-grained") observables: $\phi(\mathbf{x}') = \inf_{\mathbf{x}\setminus\mathbf{x}'} \Phi_{\boldsymbol{g}}(\mathbf{x})$ is negative entropy minimization, and $\psi(\mathbf{y}') = \Psi_{\boldsymbol{g}}\big(\mathbf{y}\big|y_{J+1} = \cdots = y_n = 0\big)$ is a projection. Under the set of partial observables $\mathbf{x}' = \boldsymbol{\nu}\mathbf{G}'$ with singular matrix $\mathbf{G}'$, the step of minimization, known as a *contraction* in the theory of probability, corresponds to taking the LFT of $\psi(\mathbf{y}') = \Psi\big(\mathbf{G}'\mathbf{y}'\big)$ (Fig. 4.1) [29].

The quantity $\phi(\mathbf{x}')$ in Fig. 4.1 is Gibbs' entropy function in classical thermodynamics,

as a function of macroscopic observables $\mathbf{x}'$. $\Psi(\boldsymbol{\mu})$ is the starting point of statistical thermodynamics when a set of $\boldsymbol{\mu}$ is in hand [10, 58, 99]. The significance of the new logic presented here is to show that a description of the energetics $\boldsymbol{\mu}$ on the level of the state space $\mathcal{S}$ can be introduced as the dual representation to statistical counting frequency $\boldsymbol{\nu}$ without evoking any prior concepts from mechanics, classical or quantum.

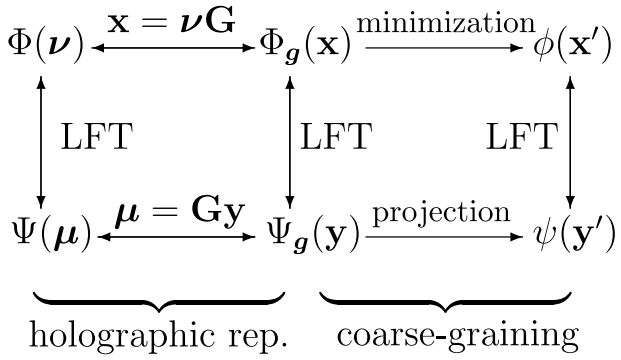

Figure 4.1: *Entropy (upper branch) and free energy (lower branch) duality via the LFT, with three different representations. Holographic* counting frequency $\boldsymbol{\nu}$ and internal energy $\boldsymbol{\mu}$; values of a complete set of observables $\mathbf{x}$ (encoded via the invertible matrix $\mathbf{G}$) and corresponding conjugate variables $\mathbf{y}$; *coarse-grained* observables with partial information $\mathbf{x}'$ and its conjugate $\mathbf{y}'$. When $\mathbf{x}' = \boldsymbol{\nu}\mathbf{G}'$ where $\mathbf{G}'$ is non-invertible, the path from $\Phi(\boldsymbol{\nu})$ to $\phi(\mathbf{x}')$ can still be taken via the lower branch: $\psi(\mathbf{y}') = \Psi(\boldsymbol{\nu}\mathbf{G}')$.

4.2.2 LFT on continuous state space, duality, and contraction

Suppose we have a system with a continuous state space $\mathcal{X}$, which (a) mathematically induces a measurable space $(\mathcal{X}, \mathcal{F})$ and (b) statistically induces a normalized empirical counting frequency distribution $\nu(y)$, where $y \in \mathcal{X}$. Let us further assume a prior probability on $\mathcal{F}$ with corresponding density function $p(y)$. Then, in parallel to (4.1a), the entropy of ν with respect to p is defined as:

$$\Phi[\boldsymbol{\nu}] = \int_{\mathcal{X}} \nu(y) \ln\left(\frac{\nu(y)}{p(y)}\right) \mathrm{d}y, \tag{4.4}$$

where the boldfaced $\boldsymbol{\nu}$ stands for the entire function $\nu(y)$ on $\mathcal{X}$. For some observable g: $\mathcal{X} \to \mathbb{R}$, the contraction of Φ to ϕ given the empirical mean value (EMV) $\overline{g}$ of g is defined

as:

$$\phi(\bar{g}) = \inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] \;\middle|\; \int_{\mathcal{X}} g(y)\nu(y)dy = \bar{g} \right\}.$$

For any bounded function $u : \mathcal{X} \to \mathbb{R}$, the free energy function corresponding to the entropy function Φ is defined as its Legendre-Fenchel transform (LFT) [110]:

$$\Psi[\mathbf{u}] := \inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] + \int_{\mathcal{X}} u(y)\nu(y)\mathrm{d}y \right\},$$

which can be obtained in closed form as [32],

$$\begin{aligned} \Psi[\mathbf{u}] &= \inf_{\boldsymbol{\nu}} \left\{ \int_{\mathcal{X}} \nu(y) \ln \left(\frac{\nu(y)}{p(y)e^{-u(y)}} \right) \mathrm{d}y \right\} \\ &= -\ln \int_{\mathcal{X}} p(y)e^{-u(y)}\mathrm{d}y. \end{aligned}$$

Furthermore, there is a duality between the contraction $\phi(\overline{g})$ and the projection $\psi(\alpha)$:

$$\begin{aligned} \psi(\alpha) &= \inf_{\overline{g}} \left\{ \phi(\overline{g}) + \alpha\overline{g} \right\} \\ &= \inf_{\overline{g}} \left\{ \inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] \;\middle|\; \int_{\mathcal{X}} g(y)\nu(y)dy = \bar{g} \right\} + \alpha\overline{g} \right\} \\ &= \inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] + \alpha \int_{\mathcal{X}} g(y)\nu(y)dy \right\} \\ &= \Psi[\alpha\boldsymbol{g}] = -\ln \int_{\mathcal{X}} e^{-\alpha g(y)} p(y)\mathrm{d}y. \end{aligned}$$

If one interprets the function $u(y)$ as internal energy, then Ψ is free energy with $k_B T = 1$. In this *energy representation*, the minimization of Φ subject to a constraint by an EMV $\overline{g}$ for an observable g is accomplished by simply substituting the internal energy function $u(y) = \alpha g(y)$, where α is the LFT conjugate variable to $\overline{g}$. In classical thermodynamics, α is interpreted as an entropic force along $\overline{g}$ with $\phi(\overline{g})$ as the potential function,

$$\alpha(\overline{g}) = -\frac{\partial\phi(\overline{g})}{\partial\overline{g}}, \tag{4.5}$$

and $\alpha(\overline{g}) \times \mathrm{d}\overline{g}$ as the corresponding "thermodynamic work."

4.2.3 Equal a priori probability on the unit ball in $\mathbb{R}^d$

Consider a point mass with random motions in $\mathbb{R}^d$; the mathematics presented in this section are general, but we are mainly interested in the cases of $d = 1, 2, 3$. Because a uniform distribution on the entire space $\mathbb{R}^d$ is not a normalizable probability, we assume that the point mass is confined to a spherical region of $\mathbb{R}^d$ with a very large radius. Mathematically, any large radius can be re-scaled to unity, so we can further assume without loss of generality that the point mass is confined to the unit ball.

Let r be the distance of the point mass from the origin within the unit ball of $\mathbb{R}^d$, and consider infinitely many i.i.d. measurements on the random variable r. Denote the EMV of r by $\overline{r}$, $0 \leq \overline{r} \leq 1$. Under the assumption of a uniform probability distribution, the logarithmic partition function, or cumulant generating function, is given by

$$\log Z_d(\lambda) = \log \int_0^1 x^{d-1} e^{-\lambda x} \mathrm{d}x + C(d), \tag{4.6}$$

in which $d > 0$, and $C(d)$ is a combinatorial constant that accounts for the d-dimensional volume element. The conjugate variable λ is determined from empirical mean value via $\xi_d(\lambda) = \overline{r}$, a sufficient statistics for λ, where the function $\xi_d(\lambda)$ is defined as [86]:

$$\xi_d(\lambda) \overset{\text{def}}{=} -\frac{\mathrm{d}}{\mathrm{d}\lambda} \log Z_d(\lambda) = \frac{\displaystyle\int_0^1 x^d e^{-\lambda x} \mathrm{d}x}{\displaystyle\int_0^1 x^{d-1} e^{-\lambda x} \mathrm{d}x}. \tag{4.7}$$

The function $\xi_d(\lambda)$ has the following essential properties:

(1) It is monotonically decreasing w.r.t. λ;

(2) $\xi_d(\lambda) \sim d\lambda^{-1}$ as $\lambda \to \infty$, $\xi_d(\lambda) \sim 1 + \lambda^{-1} + (d-1)(d+1)\lambda^{-2}$ as $\lambda \to -\infty$;

(3) Near $\lambda = 0$, $\xi_d(\lambda) = \frac{d}{d+1} - \frac{\lambda d}{(d+1)^2(d+2)} + \cdots$;

(4) For very large d, $\xi_d(\lambda)$ in Eq. 4.7 has an asymptotic expansion:

$$\xi_d(\lambda) \sim 1 - \frac{1}{d} - \frac{\lambda - 1}{d^2} + \cdots \text{ as } d \to \infty, \ \forall \text{ fixed } \lambda. \tag{4.8}$$

Fig. 4.2 shows the function $\xi_d(\lambda)$ for $d = \frac{1}{2}$, 1, and $\frac{3}{2}$. They correspond to the posterior distributions with the empirical mean value of the *squared*-distance, $x = r^2$, in D-dimensional physical space:

$$\xi_{\frac{D}{2}}(\lambda) = \frac{\int_0^1 x^{D/2} e^{-\lambda x} \mathrm{d}x}{\int_0^1 x^{D/2-1} e^{-\lambda x} \mathrm{d}x} = \frac{\int_0^1 r^{D+1} e^{-\lambda r^2} \mathrm{d}r}{\int_0^1 r^{D-1} e^{-\lambda r^2} \mathrm{d}r} , \tag{4.9}$$

with $D = 1, 2, 3$.

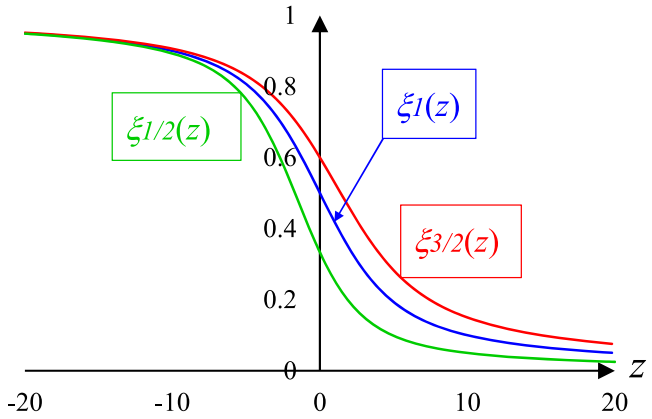

Figure 4.2: *First derivative of the cumulant generating function for the location of a point mass in the unit ball with equal* a priori *probability.*. Function $\xi_1(z) = \frac{1}{z} - \frac{e^{-z}}{1-e^{-z}}$, in blue, is monotonically decreasing with $\xi_1(0) = \frac{1}{2}$ and $\xi_1 \in [0, 1]$. It has a symmetry $\xi_1(z) + \xi_1(-z) = 1$. As $\xi_1 \to 1^-$, $z \simeq (\xi_1 - 1)^{-1} \to -\infty$. The curves $\xi_{\frac{1}{2}}(z)$ and $\xi_{\frac{3}{2}}(z)$ are shown in green and red, respectively. These three curves correspond to the posterior distribution with empirical mean value of squared-velocities in 1-D, 2-D, and 3-D (Eq. 4.9).

4.3 Generalized Gibbs' theory

In this section we formalize an assumption that is widely made in chemistry and biochemistry: applying the mathematical method in Sec. 4.2.1 to a realistic system under statistical measurements in a laboratory requires an assumption on the separation of time scales. This assumption is required to justify the statistical notion of *identically distributed* samples. A fitting example for the importance of the separation of time scales is the standard biochemical treatment of proteins as sequenced polypeptides. In reality, polypeptide chains undergo amide synthesis and hydrolysis, albeit slowly. Thus, the distinct identity of a protein only exists within a time scale $< \tau^*$ for some τ^*. On the time scale $\gg \tau^*$, a new molecular entity emerges within a chemical soup of amino acids: an ergodic "chain of random polypeptides." Studies of this latter system with the spin-glass perspective of proteins can be found in ref. [21]. On the time scale $< \tau^*$, a solution that contains different proteins is a chemical mixture; but on the time scale $\gg \tau^*$, the same solution contains a single dynamic polypeptide with stochastic sequences.

A refined understanding of the timescale τ^* emerges from the reformulation of Gibbs' theory. When combined with a set of partial observables, the principle of maximum entropy (or minimum negative entropy), a mathematical law that arises from the probability theory, partitions our description of a statistical system into a macroscopic one and a remaining random, microscopic one [29]. The former is represented by the mean values of the observables, *c.f.*, U, V, N; the latter is represented by the Gibbs distribution conditioned on the value of U, V, N. More specifically, in closed form, the parametric Gibbs distribution is expressed in terms of the conjugate variables to the observables, *c.f.*, $T^{-1}, p/T, -\mu/T$. Gibbs distribution is a posterior probability according to the modern Bayesian statistics. In R. A. Fisher's statistical theory, U, V, N are the *sufficient statistics* for the parameters T, p, μ [86, 127].

4.3.1 A hypothesis on two time scales: $t < \tau^*$ and $t \gg \tau^*$

Gibbs' chemical thermodynamics of a macromolecular ideal solution assumes that all the individual macromolecules in the solution are statistically independent and identical samples from a common probability distribution $\mathbf{p}$ on a state space $\mathcal{S}$ [31]. More importantly, each individual macromolecule is a stochastic dynamic object that fluctuates through all possible configurations in the state space over a sufficiently long time period τ^*. On this time scale, a single macromolecule can be observed to possess a stable statistical frequency distribution $\boldsymbol{\nu}$ over $\mathcal{S}$. Different individual macromolecules are statistically identical in that they possess the same $\boldsymbol{\nu}$: It is an intrinsic property of the macromolecule at this time scale. The concept of unimolecular chemical equilibrium is understood with such a time scale in mind. Under this well defined situation, one calls $\boldsymbol{\nu}^*$ the *equilibrium probability distribution* of the macromolecule, and $-\ln \boldsymbol{\nu}$ its energy landscape function [36].

On the time scale $\gg \tau^*$, $\boldsymbol{\nu}$ can become unstable, and thus different copies of the macromolecules might exhibit different frequency distributions $\boldsymbol{\nu}'$. In other words, the "fast" dynamics of the system are on $\mathcal{S}$, with a unique probability $\mathbf{p} = \boldsymbol{\nu}$, whereas the "slow" dynamics on the set of probability densities on $\mathcal{S}$, changing from one $\boldsymbol{\nu}$ to another $\boldsymbol{\nu}'$. This latter space is called the "probability simplex" on $\mathcal{S}$, denoted $\mathscr{M}_n$, and $\boldsymbol{\nu}$ and $\mathbf{p}$ are single points in $\mathscr{M}_n$ [104]. The description of a system on the time scale $\gg \tau^*$ therefore cannot be a point in $\mathcal{S}$, but instead a point in $\mathscr{M}_n$. TIn principle, the slow dynamics have a different state space than the fast dynamics.

With the separation of fast and slow time scales $t < \tau^*$ and $t \gg \tau^*$, the relation between $\mathbf{p}$ on $\mathcal{S}$ and $\Phi(\boldsymbol{\nu})$ on $\mathscr{M}_n$ becomes clear: The negative entropy function $\Phi(\boldsymbol{\nu})$ in (4.1a) provides an quantification of the probability of observing $\boldsymbol{\nu}$ on the time scale τ^* given the prior probability $\mathbf{p}$ on $\mathcal{S}$: $\ln \Pr(\boldsymbol{\nu}) = -\tau^* \Phi(\boldsymbol{\nu}) + \mathcal{O}(\tau^*)$. The scale τ^* guarantees the existence of counting frequency $\boldsymbol{\nu}$ as a stable empirical measurement. Changing $\boldsymbol{\nu}$ on the time scale $\gg \tau^*$ then is identified with the notion of *quasi-static process* in classical thermodynamics. In $\mathscr{M}_n$, the counting frequencies are ordered with respect to the prior through the theory of

entropy, and the notion of internal energy follows via the Legendre-Fenchel conjugate. The fundamental thermodynamic relation emerges from this dual relation between frequencies and energy, with conjugate variables as entropic forces.

4.3.2 From tracking to counting particles: the entropy difference

Counting the number of molecules implies neglecting distinctions between different individuals: there is an essential change in the representation of a collection of particles in mechanics and in chemistry, and this leads to a drastic difference in the computation of entropy. To illustrate this this discrepancy, let us consider N i.i.d. particles, each with $n+1$ possible discrete states with corresponding probability $\{p_k | k = 0, 1, \cdots, n\}$. Tracking means one identifies each and every particle as unique, and follows its state: $\big(k_0, k_1, \cdots, k_N\big)$. Counting, on the other hand, means that one identifies only the number of particles in each state: $\big(m_0, m_1, \cdots, m_n\big)$, where $m_0 + m_1 + \cdots + m_n = N$. We shall call these two representations of a system *Lagrangian* and *Eulerian*, respectively. We note that in the Eulerian representation, the identities of individual particles are no longer being recorded; this is a form of stereotyping according to the $n+1$ discrete states.

Because the particles are i.i.d., the Shannon entropy of the system under the Lagrangian representation is simply the Gibbs entropy, with $k_B = 1$, for a single particle with distribution $\{p_k\}$ scaled by the number of independent particles N:

$$S_{track} = -N \sum_{k=0}^{n} p_k \ln p_k \ . \tag{4.10a}$$

This quantity is on the order $\mathcal{O}(N)$ in the large N limit. On the other hand, the Shannon entropy of the system under the Eulerian representation with observed counts $\{\hat{n}_i | i = 0, 1, \cdots, n\}$ is given by:

$$S_{count} = -\sum_{\{m_i\}} \Pr\big(\{m_i\}\big) \ln \Pr\big(\{m_i\}\big), \tag{4.10b}$$

where $\Pr\big(\{m_i\}\big)$ is the probability of observing $\{\hat{n}_i\}$:

$$\Pr\big(\{m_i\}\big) := \frac{N!}{m_0!\, m_1! \cdots m_n!} p_0^{m_0} p_1^{m_1} \cdots p_n^{m_n} \ .$$

S_{count} is on the order $\mathcal{O}(n \ln N)$ in the large N limit (full derivation given in §C).

Therefore, the difference on the order of magnitude between these two entropies is entirely $S_{track} = \mathcal{O}(N)$ as $N \to \infty$, indicating that we lose significant information when we go from the Lagrangian representation to Eulerian one, regardless of the physical properties of the sub-types. Their distinction can be further illustrated as follows: When m_6, the number of molecules in state 6 among a total N molecules, changes from 5 to 4 and back to 5, the tracking can tell whether the two fives are the same or different molecules, but the counting is incapable of doing so. The difference between the two entropies reveals that the Eulerian representation neglects an individual's "life history" contained in the Lagrangian representation. In classical statistical mechanics, this difference has been recognized as *entropy of mixing*, or more perceptively *entropy of assimilation* and has led to the Gibbs' paradox [12, 90].

4.4 Continuous state space and mechanical motion

The state space for a single mechanical particle is a subset of $\mathbb{R}^6$; it is continuous. In parallel to (4.1a), the microscopic entropy function Φ for the empirical counting frequency distribution f of a single mechanical particle at $\mathbf{x} \in \mathbb{R}^3$ and with velocity $\mathbf{v} \in \mathbb{R}^3$ is the Shannon-Sanov relative entropy

$$\Phi\big[f(\mathbf{x},\mathbf{v})\big] = \iint_{\mathbb{R}^6} f(x,v) \ln\left(\frac{f(\mathbf{x},\mathbf{v})}{p(\mathbf{x},\mathbf{v})}\right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v}, \tag{4.11a}$$

where $p(x, v)$ is the assumed prior probability density function of the particle's position and velocity. In parallel to (4.1b), the corresponding free energy function Ψ is given by

$$\Psi\big[u(\mathbf{x},\mathbf{v})\big] = -\ln \iint_{\mathbb{R}^6} p(\mathbf{x},\mathbf{v}) e^{-u(\mathbf{x},\mathbf{v})} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v}. \tag{4.11b}$$

So far, we have made no reference to mechanical energy nor temperature, only to the prior probability $p(\mathbf{x}, \mathbf{v})$ and the empirical counting frequency $f(\mathbf{x}, \mathbf{v})$. These two important concepts in thermodynamics, energy and temperature, are yet to "emerge" from our statistical consideration.

In the present work, we shall assume that the $f(\mathbf{x}, \mathbf{v})$ is obtained from a large set of i.i.d. samples, idealized as data *ad infinitum*. For an ergodic mechanical system, this can be achieved either by measurements on many independent particles simultaneously or a single particle along its trajectory *ad infinitum*. The independence assumption on the samples implies an *ideal gas* and the concept of "over-damped motion." Correlations in the velocities of mechanical particles have only been carefully studied in recent years [39].

4.4.1 Conditional entropy: linear decomposition of internal energy and its additivity

We now consider a continuous mechanical state space $\mathcal{X}$ with a large number of degrees of freedom. Following the general theory outlined in ref. [105], the Shannon-Sanov entropy function is given in Eq. 4.4. According to the theory of probability, a real-valued observable is a random variable $g(y)$, $y \in \mathcal{X}$, which induces a linear functional $G[\boldsymbol{\nu}]$ as its *empirical measure value* (EMV):

$$G[\boldsymbol{\nu}] := \int_{\mathcal{X}} g(y)\nu(y)\mathrm{d}y, \tag{4.12}$$

where the boldfaced $\boldsymbol{\nu}$ stands for the entire function $\nu(y)$, $y \in \mathcal{X}$. When the observable is measured to be $\overline{g}$, this puts a constraint on the underlying frequency distribution $\nu(y)$: $G[\boldsymbol{\nu}] = \overline{g}$. Then the entropy function $\phi(\overline{g})$ and its corresponding free energy function $\psi(\alpha)$, where α is the conjugate variable to $\overline{g}$, is obtained in Sec. 4.2.2. The thermodynamics from

microscopic $\nu(y)$ to macroscopic $\overline{g}$, can be either characterized by a constrained *variational principle*:

$$\phi(\overline{g}) = \inf_{\boldsymbol{\nu}} \Big\{ \Phi[\boldsymbol{\nu}] \Big| G[\boldsymbol{\nu}] = \overline{g} \Big\}, \tag{4.13}$$

under the entropy representation, or a *conservation law* in the form of a Lagrange-Gibbs equation (see §C)

$$\mathrm{d}_{\boldsymbol{\nu}} \Phi[\boldsymbol{\nu}^*] + \alpha \, \mathrm{d}_{\boldsymbol{\nu}} G[\boldsymbol{\nu}^*] = 0, \text{ where } \alpha(\overline{g}) = -\frac{\mathrm{d}\phi}{\mathrm{d}\overline{g}}, \tag{4.14}$$

where $\boldsymbol{\nu}^*(\overline{g})$ is the minimizer of (4.13). In calculus, one solves (4.13) in terms of (4.14) with the Lagrangian multiplier α, by setting the derivatives of the Lagrangian function $\Phi[\boldsymbol{\nu}] + \alpha G[\boldsymbol{\nu}]$ w.r.t. both $\boldsymbol{\nu}$ and α to zero. The solution is a saddle point of the Lagrangian function of $\boldsymbol{\nu}$ and α. In modern convex analysis, one introduces the Lagrange dual function $\psi(\alpha)$ [110],

$$\psi(\alpha) = \inf_{\boldsymbol{\nu}} \Big\{ \Phi[\boldsymbol{\nu}] + \alpha G[\boldsymbol{\nu}] \Big\},$$

so that we can write (4.13) in terms of the *dual problem*

$$\phi(\overline{g}) = \sup_{\alpha} \Big\{ \psi(\alpha) - \alpha \overline{g} \Big\}.$$

The dual function $\psi(\alpha)$ is concave. The LFT emerges in the constrained minimization problem as originally studied by J. L. Lagrange in 1788 [22].

In addition to Eq. 4.13, the *conditional entropy* of $\nu(y)$ given the constraint $G[\boldsymbol{\nu}] = \overline{g}$ is:

$$\tilde{\Phi}[\boldsymbol{\nu} \,|\, \overline{g}\,] = \begin{cases} \Phi[\boldsymbol{\nu}] - \phi(\overline{g}) & G[\boldsymbol{\nu}] = \overline{g} \\ \infty & \text{otherwise} \end{cases} \tag{4.15}$$

That is, wherever the constraint is satisfied, the conditional entropy is finite and equal to the total entropy shifted downwards by its minimum under the constraint $G[\boldsymbol{\nu}] = \overline{g}$. Under this definition, $\tilde{\Phi}$ is non-negative and equal to zero at its minimum.

The specific $\boldsymbol{\nu}^*$ at which $\tilde{\Phi} = 0$ for the given $\overline{g}$ can be expressed in terms of the conjugate

variable α (see §C):

$$\nu^*(y;\alpha) = \frac{p(y)e^{-\alpha g(y)}}{\int_{\mathcal{X}} p(y)e^{-\alpha g(y)}\mathrm{d}y} \ . \tag{4.16}$$

For a given $\overline{g}$, α is fixed through the one-to-one relation

$$\overline{g} = \int_{\mathcal{X}} g(y)\nu^*(y;\alpha)\mathrm{d}y \ . \tag{4.17}$$

Using the identity $\Phi\big[\boldsymbol{\nu}^*\big] - \phi(\overline{g}) = 0$, we can derive an explicit expression for the rate function in (4.13), see Sec. 4.2.2:

$$\phi(\overline{g}) = \Phi[\boldsymbol{\nu}^*] = -\alpha\overline{g} + \Psi[\alpha\boldsymbol{g}]. \tag{4.18}$$

Then the LFT of $\tilde{\Phi}$ is given by (see §C):

$$\begin{aligned}
\tilde{\Psi}[\hat{\mathbf{u}}\,|\,\overline{g}] &= \inf_{\boldsymbol{\nu}} \left\{ \int_{\mathcal{X}} \nu(y)\hat{u}(y)\mathrm{d}y + \Phi[\boldsymbol{\nu}] \;\middle|\; G[\boldsymbol{\nu}] = \overline{g} \right\} - \phi\big(\overline{g}\big) && \text{(4.19a)} \\
&= \Psi[\alpha\boldsymbol{g} + \hat{\mathbf{u}}] - \alpha\overline{g} - \phi(\overline{g}) \\
&= \Psi[\alpha\boldsymbol{g} + \hat{\mathbf{u}}] - \Psi[\alpha\boldsymbol{g}] && \text{(4.19b)} \\
&= -\ln \frac{\int_{\mathcal{X}} p(y)e^{-\alpha g(y) - \hat{u}(y)}\mathrm{d}y}{\int_{\mathcal{X}} p(y)e^{-\alpha g(y)}\mathrm{d}y}, && \text{(4.19c)}
\end{aligned}$$

in which α is a function of $\hat{\mathbf{u}}$ and $\overline{g}$ through the condition

$$\overline{g} = \int_{\mathcal{X}} g(y)\hat{\nu}^*(y;\alpha)\mathrm{d}y, \quad \hat{\nu}^*(y;\alpha) = \frac{p(y)e^{-\alpha g(y) - \hat{u}(y)}}{\int_{\mathcal{X}} p(y)e^{-\alpha g(y) - \hat{u}(y)}\mathrm{d}y}. \tag{4.20}$$

Again, the constrained optimization problem in (4.19a) can be expressed as a Lagrange-Gibbs equation:

$$\hat{\mathbf{u}} \cdot \mathrm{d}\boldsymbol{\nu} + \mathrm{d}_{\boldsymbol{\nu}}\Phi[\hat{\boldsymbol{\nu}}^*] + \alpha\,\mathrm{d}_{\boldsymbol{\nu}}G[\hat{\boldsymbol{\nu}}^*] = 0, \ \alpha = -\frac{\partial\tilde{\Psi}}{\partial\overline{g}}, \tag{4.21}$$

where α and $\hat{\boldsymbol{\nu}}^*$ are functions of both $\overline{g}$ and $\hat{\mathbf{u}}$ via (4.20). We observe that when $\hat{\mathbf{u}} = 0$,

(4.21) is identical to (4.14). One can identify the three terms in (4.21) as changes in chemical energy $(\boldsymbol{\mu}/T) \cdot \mathrm{d}\mathbf{n}$, entropy $\mathrm{d}S$, and mechanical energy $T^{-1}\mathrm{d}U$, as in the Gibbs equation of standard thermodynamics.

Taken together, these results for $\tilde{\Phi}[\hat{\boldsymbol{\nu}} \,|\, \overline{g}]$ and $\tilde{\Psi}[\hat{\mathbf{u}} \,|\, \overline{g}]$ illustrate the implications of conditioning on $G[\boldsymbol{\nu}] = \overline{g}$ as follows. For statistical inference, one takes the posterior $\boldsymbol{\nu}^*$ in (4.16) as a new "prior probability" in the entropy representation. From the energetic stand point, we update the internal energy as

$$\mathbf{u} = \alpha \boldsymbol{g} + \hat{\mathbf{u}}. \tag{4.22}$$

The observable $\boldsymbol{g}$ contributes an additive term to the internal energy, with $\hat{\mathbf{u}}$ standing for the missing information, as in the case of partial observables. That is, we can write the total, holographic internal energy $\mathbf{u}$ as a linear decomposition of "the known" and "the missing."

Rewriting the equation (4.22) as $\hat{\mathbf{u}} = \mathbf{u} - \alpha \boldsymbol{g}$ is interpreted as setting $\alpha \boldsymbol{g}$ as a new reference point for the "remaining" internal energy of the system. In fact, (4.19b) shows the legitimacy of taking free energy $\Psi[\alpha \boldsymbol{g}]$ as a new reference point for the microscopic energetics $\hat{\mathbf{u}}$; and (4.19c) echos the practice of free-energy perturbation theory as an exact relation [152]. In fact, the linear constraint $G[\boldsymbol{\nu}] = \overline{g}$ on $\boldsymbol{\nu}$ corresponds to a gauge invariance on $\hat{\mathbf{u}}$, implied by (4.20):

$$\alpha(\hat{\mathbf{u}}, \overline{g}) = c \tag{4.23}$$

representing the missing information in the space of internal energies $\mathbf{u}$ (Fig. 4.3).

Fig. 4.3 illustrates the essential features of duality in Gibbs' statistical thermodynamics [105]. The spaces of $\boldsymbol{\nu}$ and $\mathbf{u}$ have been called, respectively, the *space of measures* and the *space of random variables* [60]. As outlined in Section 4.2.1, the LFT establishes a dual relation between Φ and Ψ, in which $u(y)$ is a logarithmic Radon-Nikodym derivatives between $\nu(y)$ and $p(y)$. For each observation $\boldsymbol{g}$, the relation (4.22) defines a curve $\mathbf{u} = \alpha \boldsymbol{g}$ over the conjugate variable α corresponding to $\hat{\mathbf{u}} = 0$, where α is related to the value of $\overline{g}$ through the relation in (4.17). Varying the parameter α (i.e., varying temperature) and "exploring" the entire affine space $\mathbf{u} = \alpha \boldsymbol{g}$ and its corresponding energetic representation from the point $\alpha = 0$ constitutes R. B. Laughlin's "bottom down" (Fig. 4.3) [77].

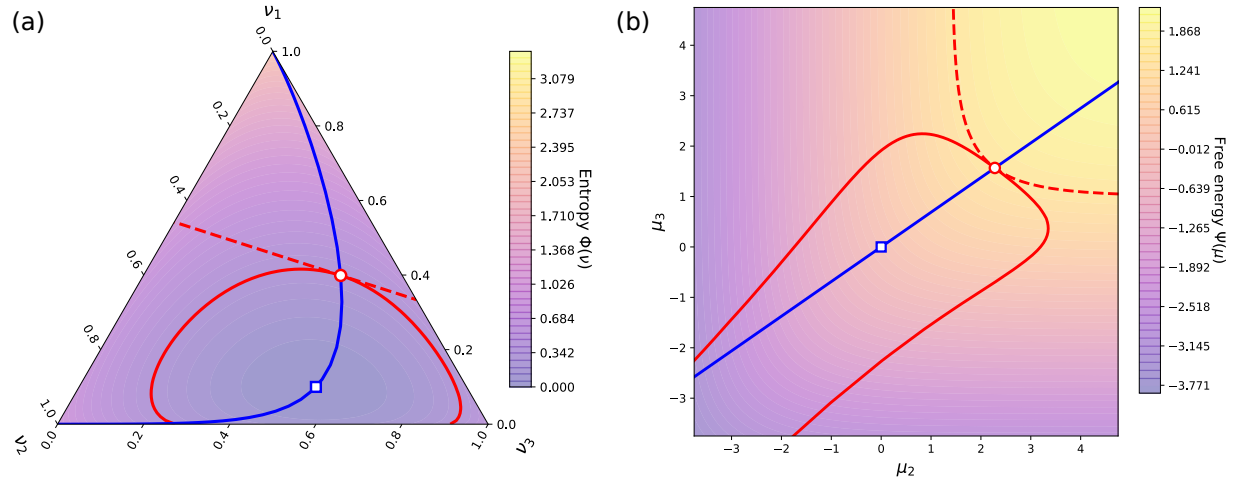

Figure 4.3: *Dual relations over a discrete state space ($n = 2$) between the (a) the entropy function Φ on space of frequencies $\boldsymbol{\nu}$ and (b) the free energy function Ψ on space of internal energies $\boldsymbol{\mu}$ (setting $\mu_1 = 0$).* The LFT establishes a one-to-one correspondence between the convex functions Φ and Ψ, and therefore, the two spaces. The blue square $\boldsymbol{\nu} = \mathbf{p}$ in (a), at which $\Phi = 0$, corresponds to the blue square $\boldsymbol{\mu} = 0$ in (b), at which $\Psi = 0$. The affine sub-space $G[\boldsymbol{\nu}] = \overline{g}$ in (a) corresponds to a gauge invariance in (b) (dashed red curves), as does the level set $\Phi[\boldsymbol{\nu}] = C$ (solid red curves). The red circle in (a) is the point at which Φ attains its minimum over $G[\boldsymbol{\nu}] = \overline{g}$; this point corresponds to the the red circle in (b), which is the point at which Ψ attains its maximum along the solid red curve in (b). The blue curve in (a) is the curve $\boldsymbol{\nu} = \boldsymbol{\nu}^*(\alpha)$ for varying α, given by (4.16), which corresponds to the blue line $\boldsymbol{\mu}^* = \alpha \boldsymbol{g}$ in (b). From this dual relationship emerges the thesis of traditional statistical thermodynamics: the dashed red line in (a) is a *microcanonical ensemble* with fixed macroscopic observable $\overline{g}$, and the dashed red curve in (b) is *Gibbs' canonical ensemble*, where α is fixed by the value of $\overline{g}$ as indicated by (4.23). At any point on the blue curve in (a), the directional derivatives of Φ and G along the dashed line are necessarily zero due to the tangent condition, whereas the the directional derivatives along the curve itself satisfy $\mathrm{d}\Phi + \alpha\, \mathrm{d}G = 0$. Therefore, each and every point on the blue curve must satisfy $\mathrm{d}_{\boldsymbol{\nu}}\Phi + \alpha\, \mathrm{d}_{\boldsymbol{\nu}}G = 0$ for all directions $\boldsymbol{\nu}$ in the simplex, *e.g.*, a Lagrange-Gibbs equation.

In the space of measures, the curve $\mathbf{u} = \alpha \boldsymbol{g}$ corresponds to the curve $\boldsymbol{\nu}^*(y;\alpha)$, as defined in (4.16) (Fig. 4.3). Considering the dual relationship between $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ in the discrete case, different corners of the simplex (i.e., "pure" states $\nu_i = 1$) in the space of measures correspond to the μ_i going to infinity in different directions in the space of random variables. Therefore, taking $\alpha \to \infty$ along the line $\boldsymbol{\mu} = \alpha \boldsymbol{g}$ corresponds to taking $\boldsymbol{\nu}$ to one of the "pure" states along the curve $\boldsymbol{\nu} = \boldsymbol{\nu}^*(\alpha)$ (Fig 4.3). On the space of all normalized densities $\boldsymbol{\nu}$, the range of $G[\boldsymbol{\nu}]$ is a compact set:

$$g_{min} \le G \le g_{max},\ g_{min} = \min_{\boldsymbol{\nu}} G[\boldsymbol{\nu}],\ g_{max} = \max_{\boldsymbol{\nu}} G[\boldsymbol{\nu}]. \tag{4.24}$$

Thus, from the standpoint of convex analysis, the information on the range of an observable G goes to "infinity" in the space of internal energies.

For a given entropy function Φ, one has a defined coordinate system in the space of $\mathbf{u}$ with a specified origin. When an observable G is given in terms of a discrete set $\{g_0, g_1, \cdots, g_n\}$, the corresponding vector $\boldsymbol{g}$ provides an alternative coordinate system over $\mathbf{u}$ at different values of α. In fact, $\boldsymbol{g}$ equips the space of measures with a "local linear structure" for the dashed red affine subspace $G[\boldsymbol{\nu}] = \overline{g}$ and blue sub-manifold $\boldsymbol{\nu} = \boldsymbol{\nu}^*(\alpha)$, with the red circle $\boldsymbol{\nu} = \boldsymbol{\nu}^*\big(\alpha(\overline{g})\big)$ as the origin (Fig. 4.3). Irrespective of the different observables, the correspondence between the $\boldsymbol{\nu}$ and $\mathbf{u}$ spaces are intrinsic to the generalized Gibbs' theory via the LFT. This intrinsic relation is the central theme of *information geometry*. In fact, the geometric "point" in the space of $\boldsymbol{\nu}$ is even more fundamental then its coordinates, as each such "point" represents a real system with the observed $\boldsymbol{\nu}$ simply being one representation of that system. Other sets of observables, *e.g.* holographic observables, are mathematically equivalent descriptions of the same system.

4.5 A theory of temperature and chemical potential of random motions

In statistical mechanics, L. Boltzmann introduced the counting frequency distribution $f(\mathbf{x}, \mathbf{v})$ for a box of ideal gas as i.i.d. particles, with $f(\mathbf{x}, \mathbf{v})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v}$ being the number of particles with

position in $[\mathbf{x}, \mathbf{x}+\mathrm{d}\mathbf{x}]$ and velocity in $[\mathbf{v}, \mathbf{v}+\mathrm{d}\mathbf{v}]$. In the infinitely long time limit, *e.g.*, at equilibrium,

$$f(\mathbf{x}, \mathbf{v}) = (N/V)e^{-\frac{U(\mathbf{x})+m|\mathbf{v}|^2/2}{k_B T}},$$

where $U(\mathbf{x})$ is a potential energy function, V is the volume of the box and N the total number of particles within. The mathematical relation between the counting frequency on the entire microscopic state space of a point mass, $\mathcal{X} = \mathbb{R}^6$, and its total internal, mechanical energy bears the name of Boltzmann-Gibbs distribution.

The Gaussian distribution for $\mathbf{v}$ was established by J. C. Maxwell. In thermodynamics, temperature T is used to characterize the random motion of particles. Here, we show that this classical concept arises from in the statistical depiction of motions following the generalized Gibbs' theory; it can be derived completely bypassing Newtonian mechanics and especially the concept of kinetic energy — the standard interpretation of temperature. We then show that the free energy of one such particle is a function of the conjugate variables for the statistical observables $|\mathbf{x}|^3$ (volume) and $|\mathbf{v}|^2$ (kinetic energy).

4.5.1 Emergence of temperature

While having conceptual significance in defining total internal, mechanical energy, a complete counting statistics and obtaining the empirical frequency $f(\mathbf{x}, \mathbf{v})$ is only a *Gedankenexperiment*. In reality, most information contained in $f(\mathbf{x}, \mathbf{v})$ is missing, and one only measures a few real-valued observables. For random mechanical motions in a homogeneous, isotropic 3-dimensional space, the most natural observable is the magnitude of velocity, $|\mathbf{v}|^2 = v_x^2 + v_y^2 + v_z^2$. The entropy function for its mean value, $\overline{|\mathbf{v}|^2}$, is obtained from maximization of the entropy in (4.11a):

$$\begin{aligned} \phi\Big(\overline{|\mathbf{v}|^2}\Big) &= \inf_{f(\mathbf{x},\mathbf{v})} \left\{ \Phi[f(\mathbf{x},\mathbf{v})] \,\middle|\, \int_{\mathbb{R}^6} |\mathbf{v}|^2 f(\mathbf{x},\mathbf{v}) \mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{v} = \overline{|\mathbf{v}|^2} \right\} \\ &= \inf_{\alpha} \Big\{ \alpha \overline{|\mathbf{v}|^2} - \psi_1(\alpha) \Big\}, \end{aligned} \tag{4.25}$$

where

$$\psi(\alpha) = -\ln \int_{\mathbb{R}^6} p(\mathbf{x}, \mathbf{v}) e^{-\alpha|\mathbf{v}|^2} \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v}, \tag{4.26}$$

and $\overline{|\mathbf{v}|^2}$ is related to α via:

$$\overline{|\mathbf{v}|^2} = \frac{\int_{\mathbb{R}^6} |\mathbf{v}|^2 p(\mathbf{x}, \mathbf{v}) e^{-\alpha|\mathbf{v}|^2} \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v}}{\int_{\mathbb{R}^6} p(\mathbf{x}, \mathbf{v}) e^{-\alpha|\mathbf{v}|^2} \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v}}. \tag{4.27}$$

Because $|\mathbf{v}|^2 \geq 0$, α must also be non-negative, as the integration of $\mathbf{v}$ in (4.27) is on the entire space $\mathbb{R}^3$. This result is the consequence of assuming a non-normalizable prior probability. When $\alpha = 0$, the mean value is expected to be in agreement with the prior probability:

$$\overline{|\mathbf{v}|^2} = \int_{\mathbb{R}^6} |\mathbf{v}|^2 p(\mathbf{x}, \mathbf{v}) \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v}.$$

On the other hand, taking $\alpha \to \infty$ corresponds to $\overline{|\mathbf{v}|^2} = 0$. Eq. 4.27 also implies that the most probable empirical frequency distribution for $\mathbf{v}$, given a measurement on the mean $\overline{|\mathbf{v}|^2}$, is

$$C(\alpha) p_v(\mathbf{v}) e^{-\alpha|\mathbf{v}|^2} \mathrm{d}\mathbf{v}, \text{ where } p_v(\mathbf{v}) = \int_{\mathbb{R}^3} p(\mathbf{x}, \mathbf{v}) \mathrm{d}\mathbf{x}, \tag{4.28}$$

α is determined by Eq. 4.27, and $C(\alpha)$ is a normalization constant.

Compared with the standard textbook, Eq. 4.28 is consistent with the assumption of *equal a priori probability* for the $p_v(\mathbf{v})$ [13, 57]. The scientific notion of temperature currently employed in physics and chemistry is gauged with respect to this choice on the prior probability. From a pure statistical standpoint without a prior knowledge from mechanics, the most significant empirical fact following the above line of thoughts is when gas particles with different masses m are mixed in a box, their empirical mean square velocity satisfies $\overline{|\mathbf{v}|^2} \propto m^{-1}$.

4.5.2 Emergence of a thermodynamic law for a particle with random motion

One characterization of an ideal gas is a collection of i.i.d. random, moving particles without internal states. With this description in mind, consider a particle with its (external) mechanical state $(\mathbf{x}, \mathbf{v}) \in \mathbb{R}^6$. We use the prior probability $p(\mathbf{x}, \mathbf{v}) = 4C\pi^{-2}|\mathbf{x}|^2|\mathbf{v}|^2$, with $|\mathbf{x}| \leq x_{max}$ and $|\mathbf{v}| \leq v_{max}$. That is, in isotropic spherical coordinates,

$$C\left(\frac{4\pi x_{max}^3}{3}\right)\left(\frac{4\pi v_{max}^3}{3}\right) = 1.$$

If we fix the center of mass at $x = 0$, identify $|\mathbf{x}|^3$ with the 3-dimensional "extent" and $|\mathbf{v}|^2$ as the "kinetic energy", then we have the free energy of one particle as a function of the conjugate variables γ and α,

$$\begin{aligned}
\mu(\alpha,\gamma) &= \ln \int_{\mathbb{R}^6} p(\mathbf{x},\mathbf{v}) e^{-\gamma|\mathbf{x}|^3 - \alpha|\mathbf{v}|^2} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v} \\
&= \ln \left\{ C \int_0^{x_{max}} 4\pi|\mathbf{x}|^2 e^{-\gamma|\mathbf{x}|^3} \mathrm{d}|\mathbf{x}| \int_0^{v_{max}} 4\pi|\mathbf{v}|^2 e^{-\alpha|\mathbf{v}|^2} \mathrm{d}|\mathbf{v}| \right\} \\
&= \ln \int_0^1 e^{-\tilde{\gamma}\tilde{x}} \mathrm{d}\tilde{x} \int_0^1 \frac{3}{2} \tilde{v}^{\frac{1}{2}} e^{-\tilde{\alpha}\tilde{v}} \mathrm{d}\tilde{v} \\
&= \ln \left\{ \frac{3}{2} \frac{(1 - e^{-\tilde{\gamma}})}{\tilde{\gamma}\tilde{\alpha}^{\frac{3}{2}}} \int_0^{\tilde{\alpha}} \hat{v}^{\frac{1}{2}} e^{-\hat{v}} \mathrm{d}\hat{v} \right\},
\end{aligned} \tag{4.29}$$

in which the non-dimensionalized variables are defined as $\tilde{\gamma} = \gamma x_{max}^3$, $\tilde{\alpha} = \alpha v_{max}^2$. If both $\tilde{\gamma}, \tilde{\alpha} \gg 1$, then Eq. 4.29 becomes

$$\mu(\alpha,\gamma) \simeq -\ln\tilde{\gamma} - \frac{3}{2}\ln\tilde{\alpha} + c, \tag{4.30}$$

where c is a constant. This is precisely the relation, according to Sackur-Tetrode's formula for ideal gas, between the chemical potential μ and $\tilde{\gamma}$, $\tilde{\alpha}$, the respective conjugate variables to volume and mean square velocity. Comparing Eq. 4.30 with the result in the standard textbook by Ben-Naim, one sees a correspondence between our $(x_{max}v_{max})^{-1}$ and $(2\pi\hbar)$ in the latter [13].

4.5.3 k_BT as a unit for "statistical mechanical energy"

Prior to any measurement, the statistical physical theory assumes a certain *equal a priori probability* on the Kolmogorovian probability space $(\Omega, \mathcal{F}, \mathbb{P})$. On the state space Ω, random variables as empirical observables are identified. Then the prior probability, under an idealized *infinitely many repeated measurements*, yields an equation among empirical mean values (EMVs) of the observables as well as a posterior canonical distribution parameterized by the EMVs. The posterior defines the concept of "thermodynamic internal energy" through conjugate variables to the empirical frequency or EMVs. The latter are sufficient statistics of the former [85, 127].

Applying this idea to the two key mechanical observables $|\mathbf{x}|^3$ and $|\mathbf{v}|^2$ in Sec. 4.5.2, with $\mathbf{x}, \mathbf{v}$ within spherical balls with radius x_{max} and v_{max}, the posterior canonical distribution then takes the form

$$\frac{|x|^2|v|^2 e^{-\gamma|\mathbf{x}|^3-\alpha|\mathbf{v}|^2}}{Z},\ (|\mathbf{x}|,|\mathbf{v}|)\in\big[0,x_{max}\big]\times\big[0,v_{max}\big]. \tag{4.31}$$

The normalization factor Z is generally a function of γ, α, x_{max} and v_{max}:

$$\begin{aligned} Z\big(\gamma,\alpha,x_{max},v_{max}\big) &= \int_0^{x_{max}} |\mathbf{x}|^2 e^{-\gamma|\mathbf{x}|^3}\mathrm{d}|\mathbf{x}| \int_0^{v_{max}} e^{-\alpha|\mathbf{v}|^2}|\mathbf{v}|^2\mathrm{d}|\mathbf{v}| \\ &= \frac{1}{6}(x_{max}v_{max})^3 \int_0^1 e^{-\tilde{\gamma}\tilde{x}}\mathrm{d}\tilde{x} \int_0^1 \tilde{v}^{\frac{1}{2}} e^{-\tilde{\alpha}\tilde{v}}\mathrm{d}\tilde{v}. \end{aligned} \tag{4.32}$$

The variables $\gamma x_{max}^3 = \tilde{\gamma}$ and $\alpha v_{max}^2 = \tilde{\alpha}$ are determined from sufficient statistics:

$$\overline{|\mathbf{x}|^3} = -\frac{\partial \ln Z}{\partial\gamma},\quad \overline{|\mathbf{v}|^2} = -\frac{\partial \ln Z}{\partial\alpha}, \tag{4.33a}$$

which yield

$$\xi_1\big(\gamma x_{max}^3\big) = \frac{\overline{|\mathbf{x}|^3}}{x_{max}^3}\ \text{ and }\ \xi_{\frac{3}{2}}\big(\alpha v_{max}^2\big) = \frac{\overline{|\mathbf{v}|^2}}{v_{max}^2}, \tag{4.33b}$$

where the monotonically decreasing functions $\xi_1(z)$ and $\xi_{\frac{3}{2}}(z)$ are shown in Fig. 4.3. Because

the assumption of finite of x_{max} and v_{max}, negative γ and α are now statistically meaningful cases. For one, $\gamma \leq 0$ when $\overline{|\mathbf{x}|^3} \geq x_{max}^3/2$. Sec. 4.2.3 further shows that when $z \to -\infty$, $\xi_{\frac{3}{2}}(z) \simeq 1 + 1/z$. Therefore,

$$\text{as } \overline{|\mathbf{v}|^2} \to v_{max}^2 \text{ from below, } \alpha \simeq \frac{1}{\overline{|\mathbf{v}|^2} - v_{max}^2}. \tag{4.34}$$

When $\overline{|\mathbf{v}|^2} \ll v_{max}^2$, the posterior velocity distribution in (4.31) is Gaussian, and the parameter α can be expressed in terms of its sufficient statistics: $\alpha^{-1} = 2\overline{|\mathbf{v}|^2}$. Comparing this with the Maxwellian distribution in the standard textbook, $\alpha \propto \frac{m}{k_B T}$. A mechanical system with smaller mean square velocity will have lower $k_B T$.

4.6 Counting i.i.d. particles in $\mathbb{R}^3$

Even when statistically independent, a collection of complex individuals can nevertheless exhibit ideal-gas-like properties in space. This is the profound insight from the chemical theory of ideal solution of very complex molecules such as proteins; it reflects the fundamental character of Poisson statistics [57]. Empirical evidence on ideal solutions goes back to the celebrated Raoult's Law in 1887 [8]. This law reveals an intriguing relation between chemical potential and partial pressure. Chemical thermodynamics beyond the ideal solution requires an extensive discussion on the choices of molarity, molality, or mole fraction as a measure for solutes [115].

The "ideal behavior" for a chemical solution is modeled after the equation of state for ideal gas $pV = Nk_B T$. Sec. 4.5.2 gives the "free energy" μ of a single mechanical particle in $\mathbb{R}^6$ in terms of the two observables $|x|^3$ and $|v|^2$. For a complex individual, μ will be similarly complicated in terms of all its internal degrees of freedom. However, as long as the spatial movements of each and every particle in the three-dimensional space are statistically independent, the solution exhibits ideal behavior.

To illustrate this concept, let us introduce the statistical problem of counting the number of particles in a simply connected observational region $\mathcal{R} \subset \mathcal{B} \subset \mathbb{R}^3$, where $\mathcal{R}$ has volume $V_{\mathcal{R}}$

and "box" $\mathcal{B}$ has volume V. For homogeneous, isotropic space, one assumes that a particle with coordinate $x \in \mathbb{R}^3$ has a uniform prior probability within $\mathcal{B}$:

$$p_x(x) = \begin{cases} \frac{1}{V} & x \in \mathcal{B} \\ 0 & x \notin \mathcal{B} \end{cases} \tag{4.35}$$

The entropy function for spatial counting is different from that of Shannon-Sanov. With increasing box size V and total number of particles N, the number of particles $n_{\mathcal{R}}$ in $\mathcal{R}$ approaches $(V_{\mathcal{R}}/V)N$:

$$\begin{aligned} \ln \Pr\{n_{\mathcal{R}} = n\} &= \ln \left\{ \frac{N!}{n!(N-n)!} \left(\frac{V_{\mathcal{R}}}{V}\right)^n \left(1 - \frac{V_{\mathcal{R}}}{V}\right)^{N-n} \right\} \\ &\simeq -n \ln n + n \ln\big(\rho V_{\mathcal{R}}\big) + n - \big(\rho V_{\mathcal{R}}\big), \end{aligned} \tag{4.36a}$$

in which we have let $N, V \to \infty$ at the same time such that $N/V = \rho$, the number density, is constant. This limit should be compared and contrasted with holding V and $V_{\mathcal{R}}$ constant while letting $N \to \infty$ with $\nu = n/N$,

$$\begin{aligned} -\ln \Pr\{n_{\mathcal{R}} = n\} &= -\ln \left\{ \frac{N!}{n!(N-n)!} \left(\frac{V_{\mathcal{R}}}{V}\right)^n \left(1 - \frac{V_{\mathcal{R}}}{V}\right)^{N-n} \right\} \\ &\simeq N \left\{ \nu \ln\left(\frac{\nu V}{V_{\mathcal{V}}}\right) + (1-\nu) \ln\left(\frac{(1-\nu)V}{(V - V_{\mathcal{V}})}\right) \right\}. \end{aligned} \tag{4.36b}$$

Eq. (4.36a) should be interpreted as the (negative) entropy being a function of volume $V_{\mathcal{R}}$, particle number n, and internal energy E, variables implicitly contained inside the number density ρ. With this identification, and $k_B = 1$, we have:

$$\frac{p}{T} = \frac{\rho}{V_{\mathcal{R}}} \left(\frac{n}{\rho} - V_{\mathcal{R}}\right), \; \frac{1}{T} = \left(V_{\mathcal{R}} - \frac{n}{\rho}\right) \frac{\partial \rho}{\partial E}, \; \Rightarrow \; p V_{\mathcal{R}} = -\left(\frac{\partial \ln \rho}{\partial E}\right)^{-1}, \tag{4.37a}$$

and the chemical potential is

$$\mu = \mu^o + T \ln \left(\frac{n}{V_{\mathcal{R}}} \right), \ \mu^o = -T \ln \rho - T. \tag{4.37b}$$

4.6.1 Counting complex i.i.d. particles

We now combine the Shannon-Sanov entropy function for the distribution of internal states in (4.4) and the entropy for counting the number of particles in an observational region $\mathcal{R}$ in 3-space in (4.36a). Let y be a point in a high dimensional space $\mathcal{X}$ for the internal degrees of freedom of a particle. Let n be the total number of i.i.d. particles in $\mathcal{R}$, with $\nu(y)$ being the frequency relative to the internal state y:

$$n \times \nu(y)\mathrm{d}y = \left\{ \begin{array}{c} \text{the \# of particles physically inside } \mathcal{R} \\ \text{and chemically in some state in } (y, y + \mathrm{d}y] \end{array} \right\}.$$

Then the (negative) entropy function is

$$n \int_{\mathcal{X}} \left(\nu(y) \ln \left(\frac{n\nu(y)}{\rho(y) V_{\mathcal{R}}} \right) \mathrm{d}y + \nu(y) - \rho(y) V_{\mathcal{R}} \right) \mathrm{d}y, \tag{4.38}$$

in which $\rho(y)$ is the expected particle density in 3-space and internal state space $\mathcal{X}$, so that $\rho(y)\mathrm{d}y\mathrm{d}x$ is the number of such particles. Eq. 4.38 can be re-written as

$$\underbrace{n \int_{\mathcal{X}} \nu(y) \ln \left(\frac{\rho_{tot} \nu(y)}{\rho(y)} \right) \mathrm{d}y}_{\text{internal entropy per particle}} \underbrace{- n \ln \frac{\rho_{tot} V_{\mathcal{R}}}{n} - n + \rho_{tot} V_{\mathcal{R}}}_{\text{Poissonian counting in 3-space}}. \tag{4.39}$$

We now introduce the LFT for the first term in (4.39) under the constraint of an arbitrary observable $g(y)$ being observed at value $\overline{g}$:

$$\int_{\mathcal{X}} g(y)\nu(y)\mathrm{d}y = \overline{g}. \tag{4.40}$$

Applying the method described in Sec. 4.2.2, the entropy function for $\overline{g}$ is specified as follows:

$$\phi(\overline{g}) = -\alpha(\overline{g})\overline{g} + \psi[\alpha(\overline{g})] = \sup_{\alpha}\{\psi(\alpha) - \alpha\overline{g}\}, \tag{4.41a}$$

$$\alpha(\overline{g}) = \arg\sup_{\alpha}\{\psi(\alpha) - \alpha\overline{g}\}, \tag{4.41b}$$

$$\psi(\alpha) = -\ln\int_{\mathcal{X}} \rho(y)e^{-\alpha g(y)}\mathrm{d}y, \tag{4.41c}$$

$$\overline{g} = \frac{\partial\psi(\alpha)}{\partial\alpha} = \frac{\int_{\mathcal{X}} g(y)\rho(y)e^{-\alpha g(y)}\mathrm{d}y}{\int_{\mathcal{X}} \rho(y)e^{-\alpha g(y)}\mathrm{d}y}. \tag{4.41d}$$

In terms of variables $G := n\overline{g}$, $V_{\mathcal{R}}$, and n, and the entropy function ϕ in (4.41a), we define the function S as

$$S(G, V_{\mathcal{R}}, n) = -n\phi\left(\frac{G}{n}\right) + n\ln\left(\frac{\rho_{tot}V_{\mathcal{R}}}{n}\right) + n - \rho_{tot}V_{\mathcal{R}}. \tag{4.42}$$

This function is Eulerian homogeneous: $S(\lambda G, \lambda V_{\mathcal{R}}, \lambda n) = \lambda S(G, V_{\mathcal{R}}, n)$. Furthermore,

$$\alpha = -\left(\frac{\partial S}{\partial G}\right)_{n, V_{\mathcal{R}}} = \frac{\partial\phi}{\partial\overline{g}}. \tag{4.43}$$

A complete parallel to the thermodynamics of ideal gas follows the homogeneous function S for any real-valued observable $g(y)$ of the internal state of an individual particle $y \in \mathcal{X}$ [23]. Velocity v is the simplest such "internal state" of a mechanical particle, and kinetic energy $\frac{1}{2}m|v|^2$ the most natural observable.

4.7 Discussion and conclusions

In the present work, the notions of "ergodic data" and "data *ad infinitum*" are used to provide a qualification of certain set of data. First, the data can be treated as samples from an identical probability distribution. Second, no matter how one starts the measurements, infinite data is expected to give exactly the same statistics.

By identifying temperature as an integration factor, Clausius' theorem established the concept of entropy and linked the empirical "energy conservation in terms of work and heat" with classical mechanical motions and mechanical energy, giving birth to a theory of random mechanical motion [41]. In the present work, starting from the theory of probability, we were able to derive the mathematical structure of classical thermodynamics, as taught in ref. [23], as well as its corresponding statistical ensemble theory. More specifically, we present the "fundamental relation" in Eqs. 4.14 and 4.21, as well as the laws for ideal gas and ideal solution in Eqs. 4.30 and 4.37. These results suggest that, under the principle of maximum entropy, any observable paired with its own conjugate variables forms a type of thermodynamic work. This result can be understood exactly as the method of Lagrange multipliers for constrained optimization [22]. Missing information, as represented by the "microscopic" quantity $\hat{\mathbf{u}} \cdot \mathrm{d}\boldsymbol{\nu}$ in Eq. 4.21, decreases with increasing linearly independent "macroscopic" observables.

While market financiers and politicians talk about "Known knowns, known unknowns, and unknown unknowns," the theory of probability makes this statement simpler and more precise: "Known, Missing Information, and Unknown" in which the middle item originates from the very conception of the Ω space in the theory of probability, *i.e.*, as a hypothesis on an imagined world, which can be fully understood when combined with data *ad infinitum*. This idea is consistent with A. Einstein's saying "Imagination is more important than knowledge." In the statistical energetics developed in the present work, the first two exhibit a mathematical additivity in terms of the concept of conjugate force and "thermodynamic work."

4.7.1 The nature of the entropy function

There is an extensive literature in applied mathematics on multiple time scales in connection to singular perturbations of dynamical systems [91]. Fixing a slow-varying variable and solving the steady state for a fast-varying variable is a routine practice when analyzing a mathematical model. The two time scales we articulate here provide both the frequentist and

Bayesian schools in statistics with a fresh outlook: our hypothesis formalizes a separation of time scales between a stochastic dynamics on $\tau^* \equiv \infty$ and its slowly varying parameters on $\gg \tau^*$. In essence, the frequentist approach ignores the $\gg \tau^*$ time scale, and thus, the existence of the prior $\mathbf{p}$. The Bayesian school believes parameters change constantly and denies the legitimacy of the time scale $\tau^* \equiv \infty$, and thus the existence of a stable frequency distribution $\boldsymbol{\nu}$. From the Bayesian perspective, a stable $\boldsymbol{\nu}$ can only exist in the mind of observers, or in a mathematical treatment of a non-autonomous dynamic model [28].

The leading condensed-matter physicist P. W. Anderson said the following in his famous paper on emergence [105]:

> *Starting with the fundamental laws and a computer, we wound have to do two impossible things — solve a problem with infinitely many bodies, and then apply the result to a finite system — before we synthesized this behavior.*[5]

What one should do if certain quantity in Anderson's "infinitely many body", macroscopic limit is zero, or infinity? For example, due to simple symmetry of isotropy, the mean velocity of a gas molecule in a box is necessarily zero. A standard approach in applied mathematics echos and extends Anderson's insight [80]: First, find a non-dimensionalized, small "ε" in the problem, then study the asymptotic limiting behavior of the problem. If the limit is zero or infinity, one should obtain the *leading order asymptotics* [91].

When the time averaged mean velocity $\overline{|v|} = 0$ in the statistical description of a gas molecule with random motions, one should obtain the observable $\overline{|v|^2}$. This observable, however, contains no information on the "shape and form" of the randomness of the particle. The modern theory of probability taught us that the Gaussian distribution that arises in this context contains no information on the probability distribution of the instantaneous v of the individual particle. Following the above approach for leading-order asymptotics, one obtains

$$\Pr\{\overline{|v|}^{\tau} = u\} \sim e^{-\tau I(u)}, \text{ where } \overline{|v|}^{\tau} = \frac{1}{\tau}\int_0^{\tau} |v(t)|\mathrm{d}t,$$

in which $I(u) \geq 0$, $u \in \mathbb{R}$, and $I(0) = 0$. The function $I(u)$ emerges as a new mathematical quantity in Anderson's limit: it is the leading order asymptotic representation of the statistics of $v(t)$. Because it is obtained as $\tau \to \infty$, $I(u)$ is a non-random quantity. Nevertheless, it captures the intrinsic randomness of the instantaneous velocity $v(t)$ of the single gas molecule. This property is exactly why and how entropy, a macroscopic quantity, represents randomness.

The mathematical limit of a sequence creates fundamental concepts that are useful to physics and chemistry. Historical examples include irrational numbers emerging from a sequence of rational numbers, instantaneous rates emerging from a sequence of infinitesimal Δx's and Δt's, and the Dirac-δ function emerging from a sequence of smooth functions. In statistical physical chemistry proper, one understands a discontinuous phase transition emerging from Gibbs' theory, which is due to non-uniform convergence of a sequence [78, 102, 146]. The concept of entropy, be that of Gibbs' or Shannon's, should be recognized and understood foremost as an asymptotic mathematical concept that emerges from a sequence of probability distributions tending towards a deterministic limit.

4.7.2 A "new" mathematical principle for statistical cells

Theoretical sciences based on mathematical principles have long been modeled after Newtonian physics: (i) A mechanistic model in the form of a set of dynamic equations is developed; (ii) mathematical analysis provides predictions; (iii) the predictions are compared with laboratory measurements. In this framework, both predictions and data are assumed to be deterministic, and any discrepancy between the two provides ideas for an improved model. This cycle of mathematical modeling has been widely adopted in practicing applied mathematics [80]. In this traditional approach, the nature of observables is not extensively discussed, neither is the method of sampling, nor is the method for comparing the mathematical prediction and empirical measurements except the choice of a "distance," usually in terms of the method of least squares.

Biophysical chemistry and single-molecule biophysics have suggested a complementary,

equally important second half to this modelling approach. Entropy, as a function of both prior probability from a stochastic model and statistics of empirical measurements *ad infinitum*, provides the necessary statistical prior on the space of all possible data. This idea is a mathematical consequence of probability, in particular the theory of large deviations [104, 105]. It encompasses the three most important relations in statistical thermodynamics proper: 1) Boltzmann's relation between energy and frequency; 2) Gibbs' fundamental thermodynamics relation in terms of observables and their conjugate variables; 3) Gibbs' statistical ensembles and transformations among them in terms of LFTs. A new principle that combines mechanistic mathematical modeling with statistical data analysis methodology emerges from this theory, as suggested schematically in Fig. 4.4.

A. N. Kolmogorov's axiomatic theory of probability forces one to explicitly state all assumptions at the onset of any probabilistic modeling. The significance of a prior assumption on probability in statistical data analysis has been emphasized by many researchers [69]. For example, as stated in the opening paragraph of ref. [67]:

> *Statistical inferences are based only in part upon the observations. An equally important base is formed by prior assumptions about the underlying situation. Even in the simplest cases, there are explicit or implicit assumptions about randomness and independence, about distributional models, perhaps prior distributions for some unknown parameters, and so on.*

The crucial distinction between the mathematical concept of probability and empirical counting frequency in statistics has not be extensively discussed in statistical mechanics literature. However, the concept of *ergodicity* already figured prominently in statistical physics, which equates the physical measure of a dynamical system with its time series sampling, for counting and for any observables.

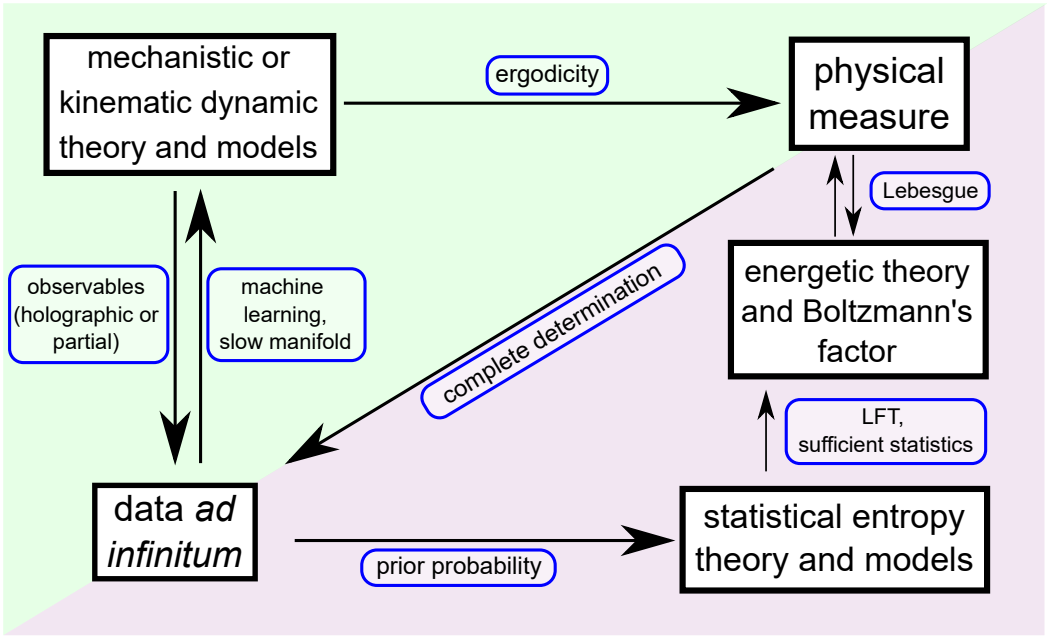

Figure 4.4: *Summary of relationship between classical mechanistic models of complex systems and new stochastic thermodynamics framework.* The upper triangle in green shows the traditional theoretical studies based on mechanistic, kinetic models, *á la* Lin and Segel [80]. Ergodicity is widely assumed in single-molecule biophysics [145]. This mathematical notion states that measurements on stationary fluctuations *ad infinitum* from an ergodic dynamical system can be used to reconstruct the steady-state probability of the system. A physical (invariant) measure represents the modern mathematical understanding for a complex dynamical system [147]. For i.i.d. sampling, the counting of all-state frequency contains holographic information [105]. Boltzmann's equilibrium $f(x, v)$ is an example, for ideal gas modelled by point masses, with the corresponding entropy function, assuming equal *a priori* probability, having the form of Shannon entropy. The entropy function for all other observables, holographic or with missing information, is mathematically derivable via the Maximum Entropy Principle, or in terms of the LFT for the constrained optimization problem. The latter leads to a projection in the energy representation (see Sec. 4.2.2 and Fig. 4.1). The lower triangle in purple encapsulates a "theory of statistical energetics" as long been understood by experimental biophysical chemists [116].

The (prior) physical measure is a consequence of an ergodic dynamics [112]. Unfortunately, for complex multi-scale dynamics, the possibility of actually obtaining such a probability from empirical measurement is null: the sample distribution among different cells is not the ergodic physical measure, even when one is confident that the latter exists. Therefore, for most statistical studies of complex systems, Kolmogorov's $\mathbb{P}$ remains a hypothesis. Using Lebesgue measure for continuous state space and counting measure for discrete state space are reasonable assumptions, but the normalization of prior probability needs to be dealt with judiciously.

Based on the mathematical principle, one concrete suggestion we have for single-cell data analysis is to identify "conjugate variables" for important biomarkers. To do this, single-cell population counting provides a statistical surrogate for the prior probability [100]. Furthermore, it could be interesting to apply the celebrated equation for chemical potential in chemistry to single-cell population dynamics, $\Delta\mu^o_{ij} = -\ln \frac{\#\text{ of cells in state } i}{\#\text{ of cells in state } j}$, and to interpret this statistical energy with respect to perturbations to a cell population. More broadly, the laboratory practice of statistical thermodynamic studies of proteins as a "procedure" (i.e., as a scientific methodology), deserves to be understood in the context of cell biology [116].

Chapter 5

QUANTIFYING SINGLE-CELL GENE EXPRESSION DYNAMICS

In the past fifteen years, advancements in cell technologies, particularly *single-cell RNA sequencing* (*scRNA-seq*), have allowed for a high depth of genomic sequencing at the resolution of the single cell [74, 83, 128]. That is, cell biologists are now able to get more and more accurate information about the genetic code and gene expression patterns of each individual cell in a population of many cells. This revolution in both the amount of and type of data being gathered has ushered in a new era of cell and systems biology [47]. In particular, more is known now about gene expression and its regulation at the level of individual cells, driving new discoveries in the realms of developmental biology and cancer biology [47, 131].

With these new data have come the development of new methods by which to analyze cell biological data. Single-cell sequencing data is particularly sparse and noisy due to technical aspects of the methodologies that generate the data [47, 48, 74, 83, 128]. As such, the data tend to undergo a large amount of pre-processing, meaning that the final reported values do not reflect exact interpretable quantities (e.g., number of mRNA molecules of a given gene in a given cell) [47, 48]. Thus, many of the methods developed to analyze scRNA-seq data have been "bioinformatic" in nature: descriptive models based in machine learning and clustering methods [87, 95, 117, 118, 119, 131, 138, 151] . More recently, however, there have been several attempts to make this analysis more rigorous and grounded in the known mechanisms of gene expression, drawing on established theories and methods from dynamical systems, stochastic processes, and biophysics [15, 37, 47, 48, 107, 114]. Our aim is to contribute to this latter "class" of single cell data analysis, specifically from the perspective outlined previously in this work: a *population dynamics* framework for cells grounded in the *stochastic*

dynamic idealization of individual cells in which a population of cell is treated as a statistical ensemble.

5.1 Direct estimation of cell state transition rates from single-cell data

Recall the linear population balance model for a multicellular population with k cell "types" (e.g., phenotypes), as laid out in §2.5:

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = \lambda_i n_i + \sum_{j=1}^{k} \nu_{ji} n_j = \sum_{j=1}^{k} a_{ij} n_j \tag{5.1}$$

where n_i denotes the number of cells of type i, the λ_i's are the net growth rates, and the ν_{ij}'s are the cell type switching rates (with $\nu_{ii} = -\sum_{j \neq i} \nu_{ij}$). In this section, we lay out a simple framework by which to estimate these parameters from single-cell transcriptomic data. This work is part of an ongoing collaboration with Sui Huang and members of the Huang Lab at the Institute for Systems Biology in Seattle, WA.

In particular, the method we derive below is applicable to experiments in which data is also available on the *lineage* (i.e., shared progenitor) of each cell. This information has classically been obtained by running parallel experiments seeded each by a single clonal population, thereby giving rise to distinct lineages [52]. Recent advances in *DNA barcoding,* however, allow for high-throughput lineage tracing in aggregated cell populations. The term DNA barcoding refers to a class of experimental methods by which a unique DNA sequence is integrated into a cells' genome (usuall by viral transfection), thus labelling all cells that descend from this "founder" cell [61]..

5.1.1 Estimation of matrix exponential entries

Recall that the general solution of (5.1) is given by $\mathbf{n}(t) = R(t)\mathbf{n}(0)$, where the vector $\mathbf{n}(0)$ represents the initial number of cells of each subtype, and the matrix $R(t)$ is given by the matrix exponential $R(t) = \exp\big(tA\big)$. Suppose that we have data $\mathbf{n}(\mathbf{0})$: the distribution of cells in each of the cell types at some reference time $t = 0$. For $t > 0$, define "lineage j"

as all living cells that descend from a progenitor cell that was of type j at $t = 0$. The jth column of the matrix $R(t)$,

$$(r_{1j}, r_{2j}, \cdots, r_{kj})^T,$$

represents the relative proportion of lineage j cells distributed in each cell type $(1, 2, \ldots, k)$ at time t.

To understand this interpretation, consider that, by definition of $R(t)$,

$$n_i(t) = \sum_{j=1}^{k} r_{ij}(t) \cdot n_j(0)\,.$$

Moreover, we an also write $n_i(t)$ as follows:

$$\begin{aligned} n_i(t) &= \sum_{j=1}^{k} (\#\text{ lineage } j \text{ cells of type } i \text{ at time } t) \\ &= \sum_{j=1}^{k} \frac{(\#\text{ lineage } j \text{ cells of type } i \text{ at time } t)}{n_j(0)} \cdot n_j(0) \end{aligned}$$

As these two expressions for $n_i(t)$ must be equal, it follows that

$$r_{ij}(t) = \frac{(\#\text{ lineage } j \text{ cells of type } i \text{ at time } t)}{n_j(0)}\,. \tag{5.2}$$

The expression in (5.2) also tells us that the sum over a given column of R gives the average per capita growth rate of that lineage:

$$\sum_{i=1}^{k} r_{ij}(t) = \frac{\#\text{ lineage } j \text{ cells of type } i \text{ at time } t}{n_j(0)} = t \cdot (\text{average growth rate of lineage } j)$$

We note that given $\mathbf{n}(0)$, the quantity $r_{ij}(t)$ in (5.2) can be directly read off from experimental data on cell types. Suppose we have this experimental data, i.e., cell type and cell lineage at times $t = 0$ and $t = T$. Denote the matrix obtained at $t = T$ from the definition

in (5.2) by $\hat{R}(T)$. Then by definition of $R(T)$,

$$\hat{R}(T) \approx R(T) = e^{TA} \implies A \approx \frac{1}{T} \log\left(\hat{R}(T)\right) =: \hat{A} \tag{5.3}$$

where 'log' above denotes the *matrix* logarithm (i.e., the inverse of the matrix exponential). By definition, the entries of A uniquely determine the cell growth and switching rates:

$$a_{ii} = \lambda_i - \sum_{j \neq i} \nu_{ij}, \quad a_i j = \nu_{ji} \ (i \neq j) \,. \tag{5.4}$$

In practice, the estimate $\hat{A}$ in (5.3) would be averaged over experiments from m possible different sampling times T_i, with each of these experiments having n_i replicates:

$$\hat{A} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n_i T_i} \sum_{\ell=1}^{n_i} \log\left(\hat{R}^{(\ell)}(T_i)\right) \tag{5.5}$$

The following is a proposed experimental-to-computational pipeline to obtain an estimate of the rate matrix A of the form (5.5):

(1) Transfect cells in culture with DNA barcodes (population of N cells will include N unique barcodes [61]); replate cells and expand cell culture.

(2) Once population is sufficiently large (e.g., $40,000$ cells), sample a fraction of cells (e.g., $35,000$ cells) for sequencing, replate and expand remaining cells ("passaging").

(3) From scRNA-seq data from sample of cell population (after pre-processing), obtain cell $\times$ gene matrix and DNA barcode label of each cell. Using this data:

   (a) Determine k cell "types" via some clustering algorithm on the data points in the cell $\times$ gene matrix (each cell = one data point); assume these correspond to the metastable states of the gene expression dynamics.

(b) From clusters and DNA barcode labels, identify "informative barcodes:" DNA barcodes of cells/cell clones that are predominantly found in one cell type. These are taken to be each "lineage j" ($j = 1, 2, \ldots, k$) as defined in the method above.

(c) Identify the fraction of cells in each of the k subtypes

(4) Fraction of cells replated and left to grow corresponds to the "initial" population distribution $\mathbf{n}(0)$. Leave cells to grow until a predetermined population size or sampling time T.

(5) Sequence cells and obtain cell $\times$ gene matrix and DNA barcode label of each cell at this time point $T > 0$. Assign "type" to each cell according to the previously determined cluster analysis and identify appropriate lineage number based on DNA barcode.

(6) Using this type $\times$ lineage for each cell, along with the plating distribution $\mathbf{n}(0)$, obtain the estimate of $R(T)$ via (5.2). Then estimate entries of A via (5.3).

(7) Average over any experimental replicates according to (5.5).

There is a subtle incompatibility between this proposed process and the estimation method described in (5.2) and (5.3): the number distribution of the initial population of replated cells over the k types, $\mathbf{n}(0)$, is necessarily not known. Furthermore, if replicates are performed by further passaging of the cell culture after Step (4), the exact number distribution at the time of sequencing, $\mathbf{n}(T)$, will also note be known, as the entire population is not represented in the sample. To overcome this problem, we assume that the sample in Step (2) is representative of the *densities* over the k cell types. That is, if we denote the relative frequencies of each type in the sequenced sample by $\mathbf{f}^{(1)}$ and that in the replated sample by $\mathbf{f}^{(2)}$, we assume that $\mathbf{f}^{(1)} = \mathbf{f}^{(2)} =: \mathbf{f}(0)$. Then we can take:

$$\mathbf{n}(0) = \mathbf{f}(0) \times \#\text{ cells plated at time } t = 0\,.$$

Similarly, if $\mathbf{f}(T)$ is the vector of relative frequencies over the k types in the sequenced sample at time T, we assume:

$$\mathbf{n}(T) = \mathbf{f}(T) \times \#\text{ cells in culture before sequencing at time } t = T\,.$$

Further experimental investigation is necessary to determine whether or not this assumption of obtaining a "representative sample" is valid.

5.1.2 Accounting for sources of noise in the data

We denote the relation in (5.3) as an approximate equality because unlike the model (5.1), the cellular processes that generate the data are not truly deterministic. Moreover, our assumptions that (1) there is a one-to-one correspondence between barcode and cell type at time $t = 0$ and (2) the samples taken for sequencing do not introduce significant bias into the cell type distributions may not always hold. There are also various technical sources of noise present within the data [48]. Finally, use of the matrix logarithm in (5.3) will introduce some level of numerical error to any estimates.

The major drawback of this relatively simple estimation framework is that it does not take into account any of these sources of error, meaning that it gives no statistical information about the parameter estimates and thus no information on how estimates can be improved. This limitation is demonstrated in §5.2, in which we apply our direct estimation method to *in silico* data generated by stochastic simulation of a population of cells. In §5.3, we discuss how our estimation method is improved upon by a recently proposed maximum likelihood estimation framework by Gunnarsson *et al.* that is applicable to the type of data we hope to be using.

5.2 Demonstration of direct estimation method on in silico *data*

In order to better understand the accuracy and other statistical properties of the method outlined in §5.1, we applied it to *in silico* data, generated to mimic the experimental data we

hope to get from an experimental setup as outlined in Steps (1)–(7). Specifically, we used the Gillespie algorithm (stochastic simulation algorithm) to simulate a continuous-time Markov chain with infinitesimal generator matrix A. That is, in a short time interval $(t, t + \Delta t)$, we assume that only one of the following events happens: a cell of some type i divides (probability $\sim b_i n_i(t)\Delta t$), a cell of some type i dies (probability $\sim d_i n_i(t)\Delta t$), or a cell of some type i switches to some other type j (probability $\sim \nu_{ij} n_i(t)\Delta t$). In this simulation, we keep track of cell type as well as barcode lineage, although the distinct barcode lineages can simply be simulated as independent realizations of this Markov process.

For this simulation, we take the number of cell types to be $k = 2$, and for each type we specify the birth (b_i) and death (d_i) rates along with the switching rates ν_{ij}. These provide the "ground truth" for the population dynamic parameters that we want to estimate, where $\lambda_i = b_i - d_i$. We also specify the initial population size and the initial relative frequencies of cells with each type and each barcode. We then simulate the process until some specified time T, and use the output of the simulation algorithm (numbers for each type $\times$ lineage combination) as our *in silico* "experimental data." Specific parameter values used for simulations are given in Table 5.1.

Table 5.1: Initial conditions, rate constants, and other simulation parameters with fixed values (units not specified). Parameters taken from Gunnarsson *et al.* (2023) [52].

Parameter	Description	Value
$n_{1,1}(0)$	number of lineage 1, type 1 cells at $t = 0$	1000
$n_{2,2}(0)$	number of lineage 2, type 2 cells at $t = 0$	1000
$n_{1,2}(0) = n_{2,1}(0)$	number of lineage i, type j $(i \neq j)$ cells at $t = 0$	0
b_1	per-capita birth rate of type 1 cells	0.6
b_2	per-capita birth rate of type 2 cells	1.0
d_1	per-capita birth rate of type 1 cells	0.3
d_2	per-capita birth rate of type 2 cells	0.5
ν_{12}	transition rate $1 \mapsto 2$	0.02
ν_{21}	transition rate $2 \mapsto 1$	0.04
T	time of data collection	4

We ran these simulations for a varying number of "experimental replicates," each time taking the average over the replicates over the estimates in accordance with (5.5) (Figure 5.1). All code used to generate these simulations can be found in the following GitHub repository: https://github.com/eeangelini/barcode_param_estimates under the folder `direct_method`. We find that the relative error of the parameter estimates scale inversely with the number of replicates, which is as expected (Figure 5.1). However, it takes a number of replicates on the order of 10^3 to achieve a relative error on the order of 10^{-3}, which is far more experimental replicates than are realistically obtainable (Figure 5.1). As we will discuss in §5.3, the maximum likelihood estimates for the same parameter values given by Gunnarsson *et al.* demonstrate a high level of accuracy with as few as three experimental replicates [52].

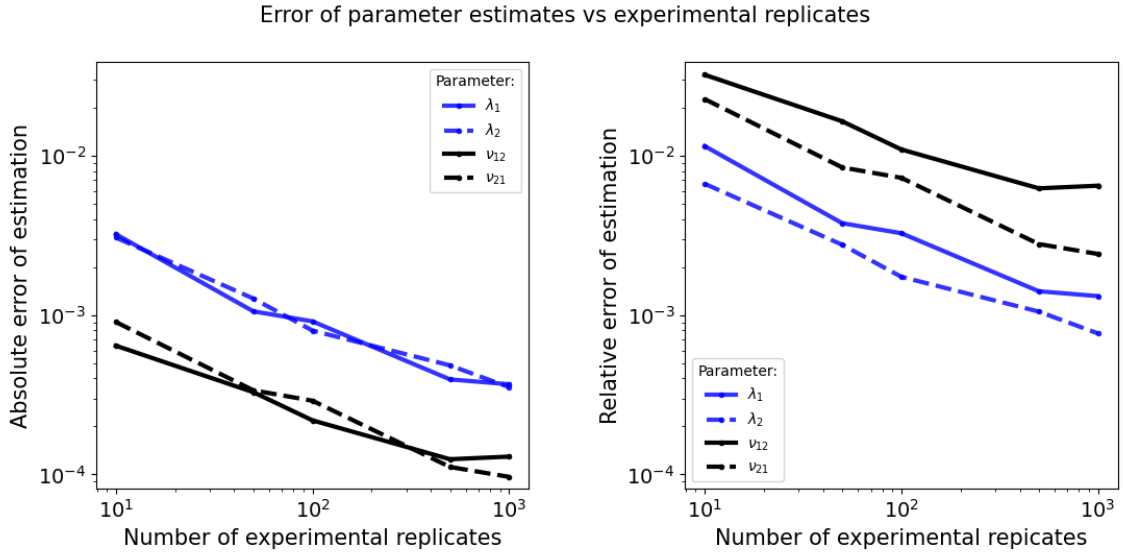

Figure 5.1: *Error of direct estimation method from §5.1 applied to* in silico *data for* $k = 2$ *cell types.* These plots show the absolute (left) and relative (right) errors of the estimates of the net growth (λ_1, λ_2) and switching rates (ν_{12}, ν_{21}) obtained from (5.5) as a function of the number of experimental replicates performed. For each fixed value of the number of replicates, estimates and errors were computed a total of 16 times, and the mean of these errors was taken and plotted above.

5.3 Adapatation of a maximum likelihood framework for estimating cell state transition and growth rates

In this section, we present the branching process model used by Gunnarsson *et al.* to describe the population dynamics of a multi-type cell population in their 2023 paper "Statistical inference of the rates of cell proliferation and phenotypic switching in cancer" [52]. We then summarize the main results derived in that publication, namely a central limit theorem for this branching process, which lends itself to a tractable maximum likelihood estimation (MLE) framework. Using this MLE method, the authors are able to find not only reliable parameter estimates for the growth and phenotype switching rates but also 95% confidence intervals for each parameter. They outline this framework for data on the number of cells of each type ("cell number data") as well as data on the relative frequency of each type ("cell fraction data"), where the data are derived from experiments in which cell populations are "founded" by mono-type colonies ("cell line experiments"). We discuss how the branching process model and resulting MLE framework is equally applicable to both cell number and cell fraction data where cell lineage is traced via DNA barcode instead of isolated cell lines.

5.3.1 Branching process model of cell growth and phenotype switching

The foundation of the statistical model presented by Gunnarsson *et al.* is a continuous time (time-homogeneous, Markovian) multitype branching process model for cell growth and phenotype switching. Suppose we have a population of cells of $K \geqslant 2$ possible types. In this population, a single type j cell divides into two type j cells at rate b_j, dies at rate d_j, and switches to type $i \neq j$ at rate ν_{ij}. Let $\mathbf{Z}^{(j)}(t)$ denote the branching process initiated by a single type j cell. That is,

$$\mathbf{Z}^{(j)}(t) = \left(Z_1^{(j)}(t)\,,\, Z_2^{(j)}(t)\,,\, \ldots\,,\, Z_K^{(j)}(t)\right)^T,$$

where $Z_i^{(j)}(t)$ denotes the number of type i descendants of the initial type j cell at time t with $Z_i^{(j)}(0) = \delta_{ij}$. The state space of $\mathbf{Z}^{(j)}(t)$ is $\Omega = \mathbb{N}_0^K$, the non-negative integer lattice in

K dimensions.

The probability distribution of this branching process is constructed via the b_i, d_i, and ν_{ij} under the assumption that individual cells are independent of each other (e.g., no cell-cell communication) [19]. This assumption of independence allows us to generalize to a process initiated by N_j cells of each type j by decomposing this process into N_j independent realizations of $\mathbf{Z}^{(j)}(t)$, $j = 1, 2, \ldots, K$ [19, 52]. We also assume that the "lifespan" of a type i cell ($i = 1, 2, \ldots, K$) is exponentially distributed with parameter $\alpha_i := b_i + d_i + \sum_{j \neq i} \nu_{ij}$ [19].[1] At the end of each type i cell's lifetime, the cell "dies" and leaves either: [19, 52]

- 0 offspring (cell death) with probability d_i/α_i,
- 2 type i offspring (cell division) with probability b_i/α_i, or
- 1 type $j \neq i$ offspring (switch types) with probability ν_{ij}/α_i.

Equivalently, in a short time interval $[t, t+\Delta t]$, a type i cell dies with probability $\sim d_i \Delta t$, divides with probability $\sim b_i \Delta t$, or switches to type $j \neq i$ with probability $\sim \nu_{ij} \Delta t$ [19, 52]. This heuristic construction of the stochastic process $\mathbf{Z}^{(j)}(t)$ reads similar to that of the birth/death/state-transition process described in §2.5 and §5.1. These two types of processes are indeed equivalent, as they are continuous time Markov chains with the same *Kolmogorov forward equation (KFE)*:

$$\frac{d}{dt} P_j(\boldsymbol{\omega}, t) = \sum_{i=1}^{K} b_i P_j(\boldsymbol{\omega} - \mathbf{e}_i, t) + d_i P_j(\boldsymbol{\omega} + \mathbf{e}_i, t) - \alpha_i P_j(\boldsymbol{\omega}, t) + \sum_{k \neq i} \nu_{ik} P_j(\boldsymbol{\omega} + \mathbf{e}_i - \mathbf{e}_k, t) \quad (5.6)$$

where $P_j(\boldsymbol{\omega}, t)$ is the transition probability semigroup for the process $\mathbf{Z}^{(j)}(t)$:

$$P_j(\boldsymbol{\omega}, t) = \Pr\left(Z^{(j)}(t) = \boldsymbol{\omega}\right), \quad \boldsymbol{\omega} \in \Omega\,.$$

[1] In fact, it is exactly this exponential distribution of the lifespan of an individual that makes this continuous time branching process Markovian [19].

The KFE is often stated in terms of the *probability generating function*:

$$F_j(\mathbf{s};t) = \mathbb{E}\left[\prod_{\ell=1}^{K} s_\ell^{Z_\ell^{(j)}(t)}\right] = \sum_{\boldsymbol{\omega}\in\Omega} P_j(\boldsymbol{\omega},t)\prod_{\ell=1}^{K} s_\ell^{\omega_\ell}\,,$$

for which the KFE in (5.6) becomes: [19, 52]

$$\frac{\partial}{\partial t}F_j(\mathbf{s};t) = \sum_{i=1}^{K}\left(b_i s_i^2 + d_i - \alpha_i s_i + \sum_{k\neq i}\nu_{ik}s_k\right)\frac{\partial}{\partial s_i}F_j(\mathbf{s};t)\,. \tag{5.7}$$

Define the *mean matrix* of the processes $\{\mathbf{Z}^{(j)}(t)\}$, $j = 1, 2, \ldots, K$ as the matrix $M(t)$ with columns $\mathbf{m}^{(j)}(t) = \mathbb{E}\big[\mathbf{Z}^{(j)}(t)\big]$:

$$m_{ij}(t) = \mathbb{E}[Z_i^{(j)}(t)]$$

It is an established result that $M(t) = \exp(tA)$, where the matrix $A \in \mathbb{R}^{K\times K}$ is the *infinitesimal generator* of $M(t)$: [19]

$$A = \begin{pmatrix} b_1 - d_1 - \sum\limits_{j\neq 1}\nu_{1j} & \nu_{21} & \cdots & \nu_{K1} \\ \nu_{12} & b_2 - d_2 - \sum\limits_{j\neq 2}\nu_{2j} & \cdots & \nu_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{1K} & \nu_{2K} & \cdots & b_2 - d_2 - \sum\limits_{j\neq K}\nu_{Kj} \end{pmatrix}. \tag{5.8}$$

Let λ^* denote the maximum eigenvalue of the matrix A, which is guaranteed to be real via the Perron-Frobenius theorem for the matrix $M(t)$ [19]. The branching process with infinitesimal parameters $\{b_i, d_i, \nu_{ij}\}$ is called *supercritical* if $\lambda^* > 0$, *critical* if $\lambda^* = 0$, and *subcritical* if $\lambda^* < 0$ [19].

Note that the infinitesimal generator A in (5.8) is exactly the matrix of rate constants in the linear population balance equation in (2.37). Moreover, the jth column of the matrix

$M(t)$ is given by:

$$\mathbb{E}\big[\mathbf{Z}^{(j)}(t)\big] = \mathbf{m}^{(j)}(t) \equiv M(t)\mathbf{e}_j = e^{tA}\mathbf{e}_j\,.$$

As previously established, the right-hand side of the above expression is the solution to (2.37) with initial condition $\mathbf{n}(0) = \mathbf{e}_j$. That is to say, the mean behavior of the branching process initiated by a single type j cell is exactly the solution to the linear population balance equation. As mentioned in §2.5, the dominant eigenvalue λ^* of A corresponds to the stationary growth rate of the overall population [148]. In terms of branching processes, this means that a supercritical process describes a population with positive net growth on average, a critical process describes one with zero net growth on average, and a subcritical process describes one with negative net growth (i.e., extinction) on average.

5.3.2 Central limit theorem for continuous time branching process

Let $\mathbf{Z}^{(j)}(t;N)$ denote the branching process initiated by N type j cells. That is, $\mathbf{Z}^{(j)}(t;N)$ is the process consisting of N i.i.d. realizations of $\mathbf{Z}^{(j)}(t)$: [52]

$$\mathbf{Z}^{(j)}(t;N) = \mathbf{Z}^{(j,1)}(t) + \mathbf{Z}^{(j,2)}(t) + \cdots + \mathbf{Z}^{(j,N)}(t)\,.$$

Gunnarsson *et al.* state and prove the following central limit theorem for $\mathbf{Z}^{(j)}(t;N)$ in the limit $N \to \infty$: [52]

$$\frac{1}{\sqrt{N}}\Big[\mathbf{Z}^{(j)}(t;N) - N\mathbf{m}^{(j)}(t)\Big] \xrightarrow{d} \mathcal{N}\big(0, \Sigma^{(j)}(t)\big)\,, \tag{5.9}$$

where $\xrightarrow{d}$ denotes convergence in distribution. The convariance matrix $\Sigma^{(j)}(t)$ is given by the expression: [52]

$$\Sigma^{(j)}(t) = 2\int_0^t M(t-\tau)\,B(\tau)\,M^T(t-\tau)\,d\tau + \operatorname{diag}\big(\mathbf{m}^{(j)}(t)\big) - \mathbf{m}^{(j)}(t)\mathbf{m}^{(j)}(t)^T\,, \tag{5.10}$$

where

$$B(\tau) = \begin{pmatrix} b_1 m_{1j}(\tau) & 0 & \cdots & 0 \\ 0 & b_2 m_{2j}(\tau) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_K m_{1K}(\tau) \end{pmatrix} = \operatorname{diag}\big(\mathbf{b} \odot \mathbf{m}^{(j)}(\tau)\big)$$

5.3.3 Statistical model of cell number data

Based on this branching process framework, Gunnarsson *et al.* propose a statistical model of experimental data that are generated as follows: [52]

- Cells in culture are sorted and plated separately by type (e.g., FACS sorting): plate j initiated (time $t = 0$) with N_j type j cells ("cell lines").
- "Cell number data" (number of cells of each type) are collected from each plate at some predetermined time point after plating, t_ℓ.
- This process of sorting, plating, and quantifying cell types is repeated for L different "end points" $t_1, t_2, \ldots, t_L$.
- For each time of data collection t_ℓ, the process is repeated a further R times to get R experimental replicates.

Let $\mathbf{n}_{j,\ell,r}$ denote the vector of cell number data collected from plate j at time t_ℓ and experimental replicate r:

$$n_{i,j,\ell,r} = \#\text{ type } i \text{ cells at time } t_\ell \text{ in population of cells initiated by } N_j \text{ type } j \text{ cells}$$

The underlying mechanistic model assumed to have generated these data is the branching process initiated by N_j cells (as described in §5.3.1–§5.3.2) sampled at $t = t_\ell$. Then if N_j is sufficiently large, the central limit theorem (5.9) gives us an approximation to the probability distribution underlying the data $\mathbf{n}_{j,\ell,r}$.

More specifically, Gunnarsson *et al.* propose the following statistical model: [52]

$$\mathbf{n}_{j,\ell,r} \sim N_j \mathbf{m}^{(j)}(t_\ell) + \mathcal{N}\big(0, N_j \Sigma^{(j)}(t_\ell)\big) + \mathcal{N}\big(0, \Sigma^{\text{err}}_{j,\ell}\big)\,. \tag{5.11}$$

The third term in this expression, the normal distribution with zero mean and covariance $\Sigma^{\text{err}}_{j,\ell}$, represents variability due to measurement error. That is, the authors assume that measurement error is Gaussian in nature, giving the following examples as possible simple models of this error: [52]

$$\Sigma^{\text{err}}_{j,\ell} = \omega^2 I \qquad \text{and} \qquad \Sigma^{\text{err}}_{j,\ell} = \omega^2 \operatorname{diag}\big(N_j \mathbf{m}^{(j)}(t_\ell)\big) \tag{5.12}$$

where ω is a parameter to be estimated from the data. In general, there may be M possible parameters $\omega_1, \omega_2, \ldots, \omega_M$ that describe the measurement error via the covariance matrix $\Sigma^{\text{err}}_{j,\ell}$. The ultimate goal of constructing this statistical model is to estimate the values of the parameters b_i, d_i, and $\nu_{i,j}$ from a given set of data from cell line experiments. Gunnarsson *et al.* do so using maximum likelihood estimation, of which we now give an overview.

5.3.4 Statement of maximum likelihood principle

Let $\boldsymbol{\theta}$ denote the vector of model parameters: the birth rates b_i, net growth rates $\lambda_i := b_i - d_i$, switching rates ν_{ij}, and measurement error parameters ω_m. These are the parameters that exactly determine the statistical model as given in (5.11) and (5.12). Thus, $\boldsymbol{\theta}$ has at least $2K + K^2$ entries (assumption of no measurement error) and up to $2K + K^2 + M$ entries, where M is the number of additional parameters used to define $\Sigma^{\text{err}}_{j,\ell}$.

According to (5.11), the probability density function for data from the jth cell line collected at time point t_ℓ (for any replicate r) is parameterized by $\boldsymbol{\theta}$ and given by:

$$f_{j,\ell}(\mathbf{n} \,|\, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^K \det\big(N_j \Sigma^{(j)}(t_\ell) + \Sigma^{\text{err}}_{j,\ell}\big)}} e^{-\frac{1}{2}\left(\mathbf{n} - N_j \mathbf{m}^{(j)}(t_\ell)\right)^T \left(N_j \Sigma^{(j)}(t_\ell) + \Sigma^{\text{err}}_{j,\ell}\right)^{-1} \left(\mathbf{n} - N_j \mathbf{m}^{(j)}(t_\ell)\right)} \tag{5.13}$$

The full set of data all independent cell lines j, replicates r, and end time points t_ℓ is given by the collection of data points

$$\mathbf{X} = \left\{ \mathbf{n}_{j,\ell,r} \,\middle|\, 1 \leqslant j \leqslant K,\, 1 \leqslant \ell \leqslant L,\, 1 \leqslant r \leqslant R \right\}.$$

These data have the following joint distribution: [52]

$$f(\mathbf{X} \,|\, \boldsymbol{\theta}) = \prod_{j=1}^{K} \prod_{\ell=1}^{L} \prod_{r=1}^{R} f_{j,\ell}(\mathbf{n}_{j,\ell,r} \,|\, \boldsymbol{\theta}). \tag{5.14}$$

The *maximum likelihood principle* states that the likelihood of a given set of parameters $\boldsymbol{\theta}$ generating the observed data is equal to the probability (density) of observing the data given parameters $\boldsymbol{\theta}$:

$$\mathcal{L}(\boldsymbol{\theta} \,|\, \mathbf{X}) = f(\mathbf{X} \,|\, \boldsymbol{\theta})$$

The *maximum likelihood estimate* of the parameters $\boldsymbol{\theta}$ from the data $\mathbf{X}$ is the value of $\boldsymbol{\theta}$ that maximizes the *likelihood function* $\mathcal{L}$:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta} \,|\, \mathbf{X}). \tag{5.15}$$

That is, given a set of observations $\mathbf{X}$ of a random variable, the most likely set of parameters that "generated" the data is one that gives the highest probability of observing $\mathbf{X}$ under the family of distributions f.

The set Θ in (5.15) is the set of "feasible" parameter values, which provides constraints for the above optimization problem. Gunnarsson *et al.* take this set to be all $\boldsymbol{\theta}$ such that $b_i \geqslant 0$, $d_i \geqslant 0$, and $\nu_{ij} \geqslant 0$ for all i, j [52]. In their implementation, Gunnarsson *et al.* find $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ by maximizing the negative double *log-likelihood function*: [52]

$$\ell(\boldsymbol{\theta} \,|\, \mathbf{X}) = -2 \ln \mathcal{L}(\boldsymbol{\theta} \,|\, \mathbf{X}). \tag{5.16}$$

That is, an equivalent expression to (5.15) for the maximum likelihood estimate is

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta} \,|\, \mathbf{X})\,. \tag{5.17}$$

Finally, a $1-\alpha$ confidence interval $[\theta_i^-, \theta_i^+]$ for each parameter θ_i is found through an optimization problem related to (5.17), derived via the likelihood ratio test: [52]

$$\theta_i^- = \min_{\boldsymbol{\theta}\in\Theta} \left\{ \theta_i \,\middle|\, \ell(\boldsymbol{\theta} \,|\, \mathbf{X}) \leqslant \ell\big(\hat{\boldsymbol{\theta}}_{\text{MLE}} \,\big|\, \mathbf{X}\big) + \chi^2_{1,1-\alpha} \right\}, \tag{5.18}$$

$$\theta_i^- = \max_{\boldsymbol{\theta}\in\Theta} \left\{ \theta_i \,\middle|\, \ell(\boldsymbol{\theta} \,|\, \mathbf{X}) \leqslant \ell\big(\hat{\boldsymbol{\theta}}_{\text{MLE}} \,\big|\, \mathbf{X}\big) + \chi^2_{1,1-\alpha} \right\}. \tag{5.19}$$

The quantity $\chi^2_{1,1-\alpha}$ denotes the $1-\alpha$ quantile of the Chi-squared distribution with one degree of freedom [52]. One interpretation of the interval $[\theta_i^-, \theta_i^+]$ is that each time we use the same process (e.g., as above) to generate this interval, the interval will contain the true value of θ_i with probability $1-\alpha$. A typical value of α to choose is $\alpha = 0.05$, i.e., 95% confidence interval.

5.3.5 Replication of results for in silico *data*

In the original manuscript, Gunnarsson *et al.* apply this MLE method to both *in silico* and existing *in vitro* datasets [52]. They implement the method in MATLAB, with all code being published on GitHub at https://github.com/egunnars/phenotypic_switching_inference/. Precise details of this implementation, including the choice of numerical optimization scheme and user-defined inputs, are given in Appendix F of Gunnarsson *et al.* [52].

We adapted this MATLAB code into Python, and applied it to the same *in silico* dataset used to generate Figure 10 in Gunnarsson *et al.* (2023). All code is available on GitHub at https://github.com/eeangelini/barcode_param_estimates under the folder `gunnarsson_mle`. This dataset was generated via the Gillespie algorithm with the same parameters as reported in Table 5.1, except for the times of data collection. Instead, these were taken to be

$$t_1 = 1, t_2 = 2, \ldots, t_6 = 6\,,$$

in accordance with the original paper [52]. Because of the "synthetic" nature of the data, we assume that there is no measurement error when building the statistical model (5.11).

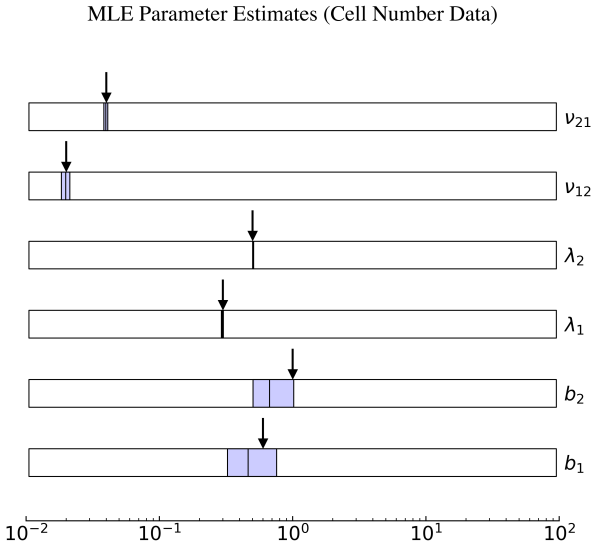

Figure 5.2: *Replication of Figure 10 from Gunnarsson* et al. *[52].* The maximum likelihood estimation framework (5.17) proposed by Gunnarsson *et al.* applied to *in silico* data for $K = 2$ cell types generated via the Gillespie algorithm. Birth, death, and switching rate parameters used to generate the data are given in Table 5.1, along with the initial cell numbers and number of experimental replicates simulated. Time points of "data collection" (end of simulation) were taken to be $t_\ell = \ell$ (units not specified). In this figure, the shaded rectangles denote the 95% confidence intervals for each parameter. The vertical line in the middle of these intervals indicates the true value of the parameter, and the arrow above the rectangle denotes the MLE.

The resulting parameter estimates are shown in Figure 5.2, which agrees with the original figure in Gunnarsson *et al.*. We note that the estimates for all parameters (net growth rates λ_i, birth rates b_i, switching rates ν_{ij}) lie within the 95% confidence intervals (Figure 5.2). Furthermore, the estimates for λ_i and ν_{ij} are relatively accurate compared to the true values used to generate the data, and the confidence intervals for these parameters are fairly small (Figure 5.2). On the other hand, the confidence intervals for the birth rates b_i are much larger, indicating a degree of uncertainty in estimating these parameters from the given data (Figure 5.2). Gunnarsson *et al.* show that this uncertainty is reduced, producing a better estimate of b_i, if data on the number of dead cells of each type is also collected [52].

5.4 Future directions

Given the nature of the single cell data we plan to obtain as described in §5.1.1, a natural next step is to implement the MLE framework proposed by Gunnarsson *et al.* for cell *fraction* data. Doing so would allow us to estimate the growth and switching rates directly from the relative frequencies of each type, instead of having to make an assumptions about unbiased sampling. The results for cell fraction are similar to that for the cell number data, as outlined in §5.3.1 - §5.3.3, except now the branching process is conditioned on survival in order to model relative frequencies: [52]

$$\mathbf{\Delta}^{(j)}(t) := \frac{\mathbf{Z}^{(j)}(t)}{\sum_{i=1}^{K} Z_i^{(j)}(t)}, \qquad \sum_{i=1}^{K} Z_i^{(j)}(t) \neq 0\,.$$

As with the MLE framework for cell number data, we would begin by translating the MATLAB code to Python, and replicating Gunnarson *et al.*'s results for *in silico* data.

Another possible way to incorporate the bias and uncertainty in the data introduced by this sampling process discussed in §5.1.1 is to incorporate it into the model of measurement error. This solution was suggested by Gunnarsson *et al.* in the original paper [52]. In general, incorporating the measurement error term into our computational implementation

of this method is something that must be done before we can use it on data generated by actual experiments.

More broadly, the applicability of the framework described in §5.3 to the data from the DNA barcoding experiments warrants further investigation. We note that the overall purpose of "cell line experiments" discussed in Gunnarsson *et al.*, and summarized here in §5.3.3, is to generate cell number data from "isolated initial conditions:" initial plating conditions in which independent lineages are founded by cells of the same type [52]. In theory, the experimental methodology of lineage tracing via DNA barcoding as laid out in §5.1.1 accomplishes this same goal (tracking of independent lineages founded each by cells of one cell type) without having to sort cells. This circumvents the technical challenge of sorting live cells by type, which requires the cell "types" in question to be identifiable in live cells (as opposed to cells that have been lysed and sequenced, i.e., transcriptomic "types"). It also simplifies the experimental setup, as K individual plates of cells do not have to be maintained in parallel, as DNA barcoding allows for lineage tracing within aggregate cell populations.

In practiced, the method by which groups of DNA barcodes are associated with distinct cell types (as described in §5.1.1) is not exact. That is to say, even if nearly all cells of type j at the time of first sequencing have DNA barcode X and vice versa, making the approximation that barcode lineage X was "founded" by only type j cells contributes to error in our parameter estimates. It is better to explicitly model this variability, if we can, rather than assuming this error is negligible. As it happens, we are able to explicitly model this spread in the initial conditions with little additional work, as Gunnarsson *et al.* give all the derivations of their results in terms of the branching process with "non-isolated" initial conditions: $\mathbf{Z}(t)$, $\mathbf{Z}(0) = \mathbf{N}$ for some vector of cell numbers $\mathbf{N}$ [52]. This is a more apt model of the data described in §5.1.1 than that of the branching process with "isolated" initial conditions described above.

One additional result we would like to obtain for this work is framing the central limit theorems derived by Gunnarsson *et al.* in terms of large deviations theory. That is, by

showing an analogous result of an asymptotic normal distribution for the branching process derived as a large deviations *ansatz* to the Kolmogorov equation (5.6) as the initial population size N tends toward infinity. In this context, the normal distribution arises from a quadratic expansion of the large deviations rate function about the mean of the branching process.

Chapter 6

CONCLUSIONS AND FUTURE DIRECTIONS

The use of Waddington's landscape as a guiding metaphor in understanding single-cell behavior in a more abstract and generalizable manner is increasingly common in the field of cell biology as a whole [122]. There is still a need, however, to not only define but also communicate a rigorous mathematical framework for a landscape that describes a cell's internal dynamics. In the present work, we have put forward one such mathematical theory, the quasi-potential landscape, and discussed the implications of this theory on the study of cell and cancer biology. More broadly, our approach to this work is grounded in the language and mathematics of biophysics and biophysical chemistry, as summarized in Figure 1.2 and Figure 6.1.

In Chapter 2, we gave an overview of the mathematical theory of a quasi-potential function for continuous, stochastic dynamics, with the motivating example in mind of gene expression dynamics in a single cell. In particular, we showed a derivation of this quasi-potential function as a large deviations rate function for a singularly perturbed stochastic differential equation. This rate function has the key property of being a Lyapunov function for the dynamics in the deterministic limit, giving it one of the desired features discussed in §1.3 that make it a "landscape" of the given system. The large deviations quasi-potential satisfies the other key feature, admitting mean transition rates between metastable states, through its relation to the theory of Freidlin and Wentzell, as discussed in §2.3.1.

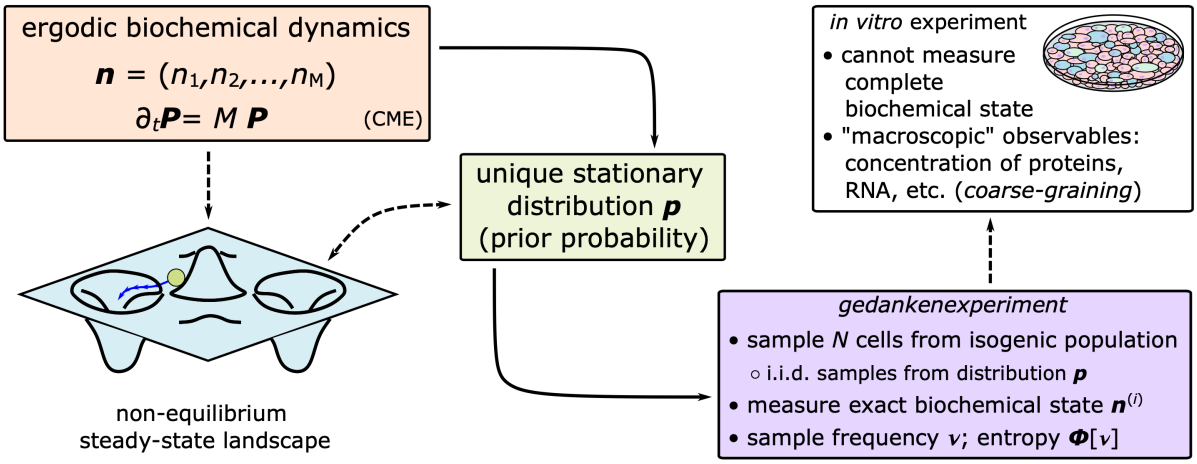

Figure 6.1: *Idealized abstraction of a single cell as an ergodic, stochastic dynamical system.* In this framework, the metastable states of the underlying biochemical dynamics are quantified and visualized as minima on a high-dimensional landscape. This landscape is, in theory, obtained as a large deviations rate function for the stationary distribution in the low-noise (e.g., many molecule) limit of the system. In an idealized thought experiment, the exact state of each cell can be measured, from which we obtain an entropy function as described in Chapter 4. In actuality, we cannot measure this exact state, as it would require measuring the number of every species of molecule in each cell in a population. Instead, we can only measure downstream indicators of this state, such as mRNA counts of a subset of genes or concentrations of certain biomarkers.

The remainder of this work builds upon two further properties of this large deviations rate function. The first is the interpretation of the rate function as a measure of *entropy* in a system, opening the theory up to the realm of classical thermodynamic analysis via the Legendre-Fenchel transform (LFT). This relation is expanded upon in Chapter 4, particularly in relation to the entropy function defined over a set of discrete counting data. In particular, Chapter 4 presents a carful treatment of the relation between microscopic states and macroscopic observables in relation to the entropy function.

The other novel framework explored in this work is the implications of the quasi-potential rate function for population dynamics, specifically those for a population of cells. We argue for a framework in which individual cells in a genetically identical ("isogenic") population are i.i.d. realizations of some underlying stochastic (biochemical, gene expression) dynamics,

making the entire population a *statistical ensemble* (Figure 1.2, Figure 6.1). At the microscopic level, each cell occupies an internal "state" (e.g., number of molecules of each type of biochemical species) with some probability determined by these dynamics (Figure 1.2, Figure 6.1). At the macroscopic level, it is not feasible to exactly measure this internal state; instead we can only observe only some downstream, coarse-grained indicator of it (Figure 1.2, Figure 6.1).

As discussed in §2.5, the quasi-potential rate function gives us a mathematical framework by which we can describe this coarse-graining, going from metastable states of the complete dynamics, to a countable number of observable states. These countable macroscopic states are the different cell "types" observed in a given isogenic cell population. The spontaneous transitions of individual cells between these types are described by the transition rates obtained from the quasi-potential. From this perspective, the overall dynamics of the population as a statistical ensemble can be described in terms of a population balance law. This population balance law can take the form of a continuous-time Markov chain, as described in §5.1, or its deterministic limit, a linear ODE, as described in §2.5 and §3.4.

BIBLIOGRAPHY

[1] R. A. Alberty. "Legendre transforms in chemical thermodynamics." *Chem. Rev.* **94**.6 (1994), pp. 1457–1482.

[2] S. Altschuler and L. Wu. "Cellular heterogeneity: do differences make a difference?" *Cell* **141** (2010), pp. 559–563.

[3] A. Anderson and V. Quaranta. "Integrative mathematical oncology." *Nat. Rev.* **8** (2008), pp. 227–234.

[4] A. Anderson *et al.* "Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment." *Cell* **127** (2006), pp. 905–915.

[5] P. W. Anderson. "More is different: Broken symmetry and the nature of the hierarchical structure of science." *Science* **177** (1972), pp. 393–396.

[6] E. Angelini *et al.* "A model for the intrinsic limit of cancer therapy: duality of treatment-induced cell death and treatment-induced stemness." *PLOS Comput. Biol.* **18** (2022), e1010319.

[7] P. Ao. "Laws in Darwinian evolutionary theory." *Phys. Life Rev.* **2**.2 (2005), pp. 117–156. ISSN: 1571-0645.

[8] P. W. Atkins. *Physical Chemistry*. Oxford, U.K.: Oxford University Press, 1978.

[9] C. S.-O. Attolini and F. Michor. "Evolutionary theory of cancer." *Ann N.Y. Acad. Sci.* **1168** (2009), pp. 23–51.

[10] D. E. Barrick. *Biomolecular Thermodynamics: From Theory to Application*. Boca Raton: CRC Press, 2018.

[11] R. A. Beckman, G. Schemmann, and C.-H. Yeang. "Impact of genetic dynamics and single-cell heterogeneity on development of nonstandard personalized medicine strategies for cancer." *Proc. Natl. Acad. Sci. U.S.A.* **109** (2012), pp. 14586–14591.

[12] A. Ben-Naim. "Entropy of mixing and entropy of assimilation, an information-theoretical perspective." *Am. J. Phys.* **74** (2006), pp. 1126–1135.

[13] A. Ben-Naim. *Statistical Thermodynamics for Chemists and Biochemists*. New York: Springer, 1992.

[14] A.-M. Bleau *et al.* "Pten/pi3k/akt pathway regulates the side population phenotype and abcg2 activity in glioma tumor stem-like cells." *Cell Stem Cell* **4** (2009), pp. 226–235.

[15] F. Bocci *et al.* "Theoretical and computational tools to model multistable gene regulatory networks." *Rep. Prog. Phys.* **86** (2023), p. 106601.

[16] S. Bornholdt and S. Kauffman. "Ensembles, dynamics, and cell types: revisiting the statistical mechanics perspective on cellular regulation." *J. Theor. Biol.* **467** (2019), pp. 15–22.

[17] I. Bozic. "Quantification of the selective advantage of driver mutations is dependent on the underlying model and stage of tumor evolution." *Cancer Res.* **82** (2022), pp. 21–24.

[18] R. Brady-Nicholls *et al.* "Prostate-specific antigen dynamics predict individual responses to intermittent androgen deprivation." *Nat. Commun.* **11** (2020).

[19] *Branching Processes*. Mineola, New York: Dover Publications, Inc., 1972.

[20] A. Brock, H. Chang, and S. Huang. "Non-genetic heterogeneity – a mutation-independent driving force for the somatic evolution of tumors." *Nat. Rev. Genet.* **10** (2009), pp. 336–342.

[21] J. D. Bryngelson and P. G. Wolynes. "Spin glasses and the statistical mechanics of protein folding." *Proc. Natl. Acad. Sci. U. S. A.* **84**.21 (1987), pp. 7524–7528.

[22] P. Bussotti. "On the genesis of the Lagrange multipliers." *J. Optim. Theory Appl.* **117** (2003), pp. 453–459.

[23] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. 2nd. New York: Wiley, 1991.

[24] C. Chaffer *et al.* "Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state." *Proc. Natl. Acad. Sci. U.S.A.* **108** (2011), pp. 7950–7955.

[25] S. Corrêa *et al.* "Wnt/β-catenin pathway regulates abcb1 transcription in chronic myeloid leukemia." *BMC Cancer* **12** (2012).

[26] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. 2nd. New York: Wiley-Interscience, 2006.

[27] D. R. Cox. "Regression models and life-tables." *J. R. Stat. Soc.* **34** (1972), pp. 187–220.

[28] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences." *Phys. Rev. E* **60** (3 1999), pp. 2721–2726.

[29] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. 2nd ed. New York: Springer, 1998.

[30] L. A. Diaz Jr *et al.* "The molecular evolution of acquired resistance to targeted egfr blockade in colorectal cancers." *Nature* **486** (2012), pp. 537–540.

[31] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. New York: Oxford University Press, 1986.

[32] D. V. Donsker and S. R. S. Varadhan. "Asymptotic evaluation of certain Markov process expectations for large time, IV." *Commun. Pure Appl. Math.* **36**.2 (1983), pp. 183–212.

[33] S. W. Englander and L. Mayne. "The case for defined protein folding pathways." *Proc. Natl. Acad. Sci. U.S.A.* **114** (2017), pp. 8253–8258.

[34] F. Fang *et al.* "Genomic and epigenomic signatures in ovarian cancer associated with resensitization to platinum drugs." *Cancer Res.* **78** (2018), pp. 631–644.

[35] E. F. Fornasiero *et al.* "Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions." *Nat. Commun.* **9** (2018), p. 4230.

[36] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. "The energy landscapes and motions of proteins." *Science* **254** (1991), pp. 1598–1603.

[37] S. L. Freedman *et al.* "A dynamical systems treatment of transcriptomic trajectories in hematopoiesis." *Development* **150** (2023), dev201280.

[38] M. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. third. New York: Springer-Verlag, 2012. ISBN: 9783642258466.

[39] W. Fu, Y. Zhang, and H. Zhao. "Universal scaling of the thermalization time in one-dimensional lattices." *Phys. Rev. E* **100** (1 2019), p. 010101.

[40] J. Gallaher *et al.* "Spatial heterogeneity and evolutionary dynamics modulate time to recurrence in continuous and adaptive cancer therapies." *Cancer Res.* **78** (2018), pp. 2127–2139.

[41] G. Gallavotti. *Statistical Mechanics: A Short Treatise*. New York: Springer, 1999.

[42] H. Ge and H. Qian. "Analytical mechanics in stochastic dynamics: most probable path, large deviation rate function, and hamilton-jacobi equation." *Int. J. Mod. Phys. B* **20** (2012).

[43] H. Ge and H. Qian. "Featured review: random perturbations of dynamical systems. third edition." *SIAM Rev.* **55** (2012).

[44] D. Gillies. *Philosophical Theories of Probability*. London, U.K.: Routledge, 2000.

[45] M. Gluzman, J. Scott, and A. Vladimirsky. "Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory." *Proc. R. Soc. B.* **287** (2020).

[46] L. González-Silva, L. Quevedo, and I. Varela. "Tumor functional heterogeneity unraveled by scrna-seq technologies." *Trends in Cancer* **6** (2020), pp. 13–19.

[47] G. Gorin, J. J. Vastola, and L. Pachter. "Studying stochastic systems biology of the cell with single-cell genomics data." *Cell Systems* **14** (2023), pp. 822–843.

[48] G. Gorin *et al.* "Rna velocity unraveled." *PLOS Comput. Biol.* **18**.9 (2022), e1010492.

[49] M. Greaves and C. C. Maley. "Clonal evolution in cancer." *Nature* **481** (2012), pp. 306–313.

[50] J. Greene, C. Sanchez-Tapia, and E. Sontag. "Mathematical details on a cancer resistance model." *Front. Bioeng. Biotechnol.* **8** (2020).

[51] E. A. Guggenheim. *Modern Thermodynamics by the Methods of Willard Gibbs*. New York: Methuen & Co., 1933.

[52] E. Gunnarsson, J. Foo, and K. Leder. "Statistical inference of the rates of cell proliferation and phenotypic switching in cancer." *J. Theor. Biol.* **568** (2023), p. 111497.

[53] P. Gupta *et al.* "Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells." *Cell* **146** (2011), pp. 633–644.

[54] P. Hahnfeldt and L. Hlatky. "Cell resensitization during protracted dosing of heterogeneous cell populations." *Radiat. Res.* **150** (1998), pp. 681–687.

[55] D. Hanahan and R. A. Weinberg. "The hallmarks of cancer." *Cell* **100** (2000), pp. 57–70.

[56] D. Hanahan and R. A. Weinberg. "The hallmarks of cancer: the next generation." *Cell* **144** (2011), pp. 646–669.

[57] T. L. Hill. *Statistical Mechanics: Principles and Selected Applications*. New York: McGraw-Hill, 1956.

[58] T. L. Hill. *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*. New York: Springer-Verlag, 1985.

[59] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. first. Cambridge: Cambridge University Press, 1998. ISBN: 052162570X.

[60] L. Hong, H. Qian, and L. F. Thompson. "Representations and divergences in the space of probability measures and stochastic thermodynamics." *J. Comput. Appl. Math.* **376** (2020), p. 112842.

[61] K. K. Howland and A. Brock. "Cellular barcoding tracks heterogeneous clones through selective pressures and phenotypic transitions." *Trends in Cancer* **9** (2023), pp. 591–601.

[62] S. Huang. "Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells." *Cancer Metastasis Rev.* **32** (2013), pp. 421–448.

[63] S. Huang. "The logic of cancer treatment: why is it so hard to cure cancer treatment-induced progression." In: *Rethinking Cancer: A New Paradigm for the Post-genomics Era*. Ed. by B. Strauss *et al.* Cambridge, MA: MIT Press, 2021, pp. 63–128.

[64] S. Huang, I. Ernberg, and S. Kauffman. "Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective." *Semin. Cell. Dev. Biol.* **20** (2009), pp. 869–876.

[65] S. Huang *et al.* "Cell fates as high-dimensional attractor states of a complex gene regulatory network." *Phys. Rev. Lett.* **94** (2005).

[66] S. Huang *et al.* "Processes on the emergent landscapes of biochemical reaction networks and heterogeneous cell population dynamics: Differentiation in living matters." *J. R. Soc. Interface* **14** (2017), p. 20170097.

[67] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. 2nd. New York: John Wiley & Sons, 2009.

[68] T.-H. Hung *et al.* "Wnt5a regulates abcb1 expression in multidrug-resistant cancer cells through activation of the non-canonical pka/β-catenin pathway." *Oncotarget* **5** (2014), pp. 12273–12290.

[69] E. T. Jaynes. *Probability Theory: The Logic of Science*. London, U.K.: Cambridge University Press, 2003.

[70] D.-Q. Jiang, Y. Wang, and D. Zhou. "Phenotypic equilibrium as probabilistic convergence in multi-phenotype cell population dynamics." *PLOS One* **12** (2017).

[71] S. Kauffman. "Differentiation of malignant to benign cells." *J. Theor. Biol* **31** (1971), pp. 429–451.

[72] K. Kemper *et al.* "Phenotype switching: tumor cell plasticity as a resistance mechanism and target for therapy." *Cancer Res.* **74** (2014), pp. 5937–5941.

[73] M. Kimmel, D. E. Axelrod, and G. M. Wahl. "A branching process model of gene amplification following chromosome breakage." *Mutat. Res.* **276** (1992), pp. 225–239.

[74] A. M. Klein *et al.* "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." *Cell* **161** (2015), pp. 1187–1201.

[75] E. Kozlowska *et al.* "Mathematical modeling predicts response to chemotherapy and drug combinations in ovarian cancer." *Cancer Res.* **78** (2018), pp. 4036–4044.

[76] E. Kozlowska *et al.* "Mathematical modeling predicts response to chemotherapy in advanced non-resectable non-small cell lung cancer patients treated with platinum-based doublet." *PLOS Comput. Biol.* **16** (2020).

[77] R. B. Laughlin. *A Different Universe: Reinventing Physics from Bottom Down*. Cambridge: Basic Books, 2005.

[78] T. D. Lee and C. N. Yang. "Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model." *Phys. Rev.* **87** (1952), pp. 410–419.

[79] Q. Li *et al.* "Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape." *Proc. Natl. Acad. Sci. U.S.A.* **113** (2016), pp. 2672–2677.

[80] C.-C. Lin and L. A. Segel. *Mathematics Applied to Deterministic Problems in the Natural Sciences*. Philadelphia: SIAM, 1988.

[81] T. Liu *et al.* "Exosome-transmitted mir-128-3p increase chemosensitivity of oxaliplatin-resistant colorectal cancer." *Mol. Cancer* **18** (2019).

[82] Z. Lu and H. Qian. "Emergence and breaking of duality symmetry in thermodynamic behavior: Repeated measurements and macroscopic limit." *Phys. Rev. Lett.* **128** (2022), p. 150603.

[83] E. Z. Macosko *et al.* "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets." *Cell* **161** (2015), pp. 1202–1214.

[84] S. Maetani and J. M. Gamel. "Evolution of cancer survival." *Surg. Oncol.* **19** (2010), pp. 49–51.

[85] B. Mandelbrot. "The role of sufficiency and of estimation in thermodynamics." *Ann. Math. Stat.* **33** (1962), pp. 1021–1038.

[86] B. Mandelbrot. "On the derivation of statistical thermodynamics from purely phenomenological principles." *J. Math. Phys.* **5**.2 (1964), pp. 164–171.

[87] N. Marjanovic *et al.* "Emergence of a high-plasticity cell state during lung cancer evolution." *Cancer Cell* **38** (2020), pp. 229–246.

[88] C. Neftel *et al.* "An integrative model of cellular states, plasticity, and genetics for glioblastoma." *Cell* **178** (2019), pp. 835–849.

[89] B. Nolting and K. Abbott. "Balls, cups, and quasi-potentials: quantifying stability in stochastic systems." *Ecology* **97** (2016), pp. 850–864.

[90] R. M. Noyes. "Entropy of mixing of interconvertible species: Some reflections on the Gibbs paradox." *J. Chem. Phys.* **34**.6 (1961), pp. 1983–1985.

[91] R. E. O'Malley. *Historical Developments in Singular Perturbations*. Cham, Switzerland: Springer, 2014.

[92] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. "Theory of protein folding: the energy landscape perspective." *Ann. Rev. Phys. Chem.* **48**.1 (1997), pp. 545–600.

[93] Y. Oren *et al.* "Cycling cancer persister cells arise from lineages with distinct programs." *Nature* **596** (2021), pp. 576–582.

[94] J. Pacheco, F. Santos, and D. Dingli. "The ecology of cancer from an evolutionary game theory perspective." *Interface Focus* **4** (2014).

[95] S. Persad *et al.* "Seacells infers transcriptional and epigenomic cellular states from single-cell genomics data." *Nat. Biotechnol.* **41** (2023), pp. 1746–1757.

[96] S. Piana, J. L. Klepeis, and D. E. Shaw. "Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations." *Curr. Opin. Struct. Biol.* **24** (2014), pp. 98–105.

[97] A. Pisco and S. Huang. "Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'what does not kill me strengthens me'." *Br. J. Cancer* **112** (2015), pp. 1725–1732.

[98] A. Pisco *et al.* "Non-darwinian dynamics in therapy-induced cancer drug resistance." *Nat. Commun.* **4** (2013).

[99] D. Poland and H. Scheraga. *Theory of Helix-Coil Transitions in Biopolymers: Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules.* New York: Academic Press, 1970.

[100] H. Qian and Y.-C. Cheng. "Counting single cells and computing their heterogeneity: from phenotypic frequencies to mean value of a quantitative biomarker." *Quant. Biol.* **8** (2020), pp. 172–176.

[101] H. Qian and H. Ge. *Stochastic Chemical Reaction Systems in Biology.* Cham, Switzerland: Springer Nature, 2021.

[102] H. Qian *et al.* "A framework towards understanding mesoscopic phenomena: Emergent unpredictability, symmetry breaking and dynamics across scales." *Chem. Phys. Lett.* **665** (2016), pp. 153–161.

[103] H. Qian. "From discrete protein kinetics to continuous brownian dynamics: a new perspective." *Protein Sci.* **11** (2002), pp. 1–5.

[104] H. Qian. "Internal energy, fundamental thermodynamic relation, and Gibbs' ensemble theory as laws of statistical counting" (2022). arXiv: 2209.14337.

[105] H. Qian. "Statistical chemical thermodynamics and energetic behavior of counting: Gibbs' theory revisited." *J. Chem. Theory Comput.* **18**.11 (2022), pp. 6421–6436.

[106] H. Qian. "Stochastic physics, complex systems and biology." *Quant. Biol.* **1** (2013), pp. 50–53.

[107] X. Qiu *et al.* "Mapping transcriptomic vector fields of single cells." *Cell* **185** (2022), pp. 690–711.

[108] A. Radzicka and R. Wolfenden. "Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases." *J. Am. Chem. Soc.* **118**.26 (1996), pp. 6105–6109.

[109] A. Read, T. Day, and S. Huijben. "The evolution of drug resistance and the curious orthodoxy of aggressive chemotherapy." *Proc. Natl. Acad. Sci. U.S.A.* **108** (2011), pp. 10871–10877.

[110] R. T. Rockafellar. *Convex Analysis*. Princeton: Princeton University Press, 1970.

[111] A. Rosech *et al.* "A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth." *Cell* **141** (2010), pp. 583–594.

[112] D. Ruelle. *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*. 2nd. Cambridge, U.K.: Cambridge University Press, 2004.

[113] E. D. Saad and A. Katz. "Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometime loosely defined." *Ann. Oncol.* **20** (2009), pp. 460–464.

[114] M. Sáez, J. Briscoe, and D. Rand. "Dynamical landscapes of cell fate decisions." *Interface Focus* **12** (2022), p. 20220002.

[115] J. A. Schellman. "Fifty years of solvent denaturation." *Biophys. Chem.* **96**.2-3 (2002), pp. 91–101.

[116] J. A. Schellman. "Thermodynamics, molecules and the Gibbs conference." *Biophys. Chem.* **64**.1 (1997), pp. 7–13.

[117] G. Schiebinger *et al.* "Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming." *Cell* **176** (2019), pp. 928–943.

[118] M. Setty *et al.* "Characterization of cell fate probabilities in single-cell data with palantir." *Nat. Biotechnol.* **37** (2019), pp. 451–460.

[119] Y. Sha *et al.* "Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data." *Nucleic Acids Res.* **48** (2020), pp. 9505–9520.

[120] S. Shaffer *et al.* "Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors." *Cell* **182** (2020), pp. 947–959.

[121] S. Shaffer *et al.* "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance." *Nature* **546** (2017), pp. 431–435.

[122] N. Shakiba *et al.* "Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors." *Cell Systems* **13** (2022), pp. 4–9.

[123] A. Sharma *et al.* "Longitudinal single-cell rna sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy." *Nat. Commun.* **9** (2018).

[124] M. Strobl *et al.* "Turnover modulates the need for a cost of resistance in adaptive therapy." *Cancer Res.* **81** (2021), pp. 1135–1147.

[125] Y. Su *et al.* "Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance." *Proc. Natl. Acad. Sci. U.S.A.* **114** (2017), pp. 13679–13684.

[126] M. Suva and I. Tirosh. "Single-cell rna sequencing in cancer: lessons learned an emerging challenges." *Mol. Cell* **75** (2019), pp. 7–12.

[127] L. Szilard. "Über die ausdehnung der phänomenologschen thermodynamik auf die schwankungserscheinungen." *Zeitschrift für Physik* **32** (1925), pp. 753–7888.

[128] F. Tang *et al.* "Mrna-seq whole-transcriptome analysis of a single cell." *Nat. Methods* **6** (2009), pp. 377–382.

[129] M. Tong *et al.* "Transcriptomic but not genomic variability confers phenotype of breast cancer stem cells." *Cancer Commun.* **38** (2018), pp. 1–16.

[130] A. Tran *et al.* "Delicate balances in cancer chemotherapy: modeling immune recruitment and emergence of systemic drug resistance." *Front. Immunol.* **11** (2020).

[131] C. Trapnell *et al.* "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells." *Nat. Biotechnol.* **32** (2014), pp. 381–386.

[132] N. Vasan, J. Baselga, and D. Hyman. "A view on drug resistance in cancer." *Nature* **575** (2019), pp. 299–309.

[133] J. Wang. "Landscape and flux theory of non-equilibrium dynamical systems with application to biology." *Adv. Phys.* **64**.1 (2015), pp. 1–137.

[134] J. Wang, L. Xu, and E. Wang. "Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations." *Proc. Natl. Acad. Sci. U.S.A.* **105** (2008), pp. 12271–12276.

[135] J. Wang *et al.* "Quantifying the waddington landscape and biological paths for development and differentiation." *Proc. Natl. Acad. Sci. U.S.A.* **108** (2011), pp. 8257–8262.

[136] J. Wang *et al.* "The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation." *Biophys. J.* **99** (2010), pp. 29–39.

[137] S. Wang *et al.* "Gene amplification and multidrug resistance induced by the phosphatase-inhibitory tumor promoter, okadaic acid." *Carcinogenesis* **16** (1995), pp. 637–641.

[138] Z. Wang *et al.* "Markovhc: markov hierarchical clustering for the topological structure of high-dimensional single-cell omics data with transition pathway and critical point detection." *Nucleic Acids Res.* **50** (2022), pp. 46–56.

[139] R. Weinstock. *Calculus of Variations*. first. New York: Dover Publications, Inc., 1974. ISBN: 9780486630694.

[140] C.-H. Weng *et al.* "Epithelial-mesenchymal transition (emt) beyond egfr mutations per se is a common mechanism for acquired resistance to egfr tki." *Oncogene* **38** (2018), pp. 455–468.

[141] J. West, Y. Ma, and P. K. Newton. "Capitalizing on competition: an evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment." *J. Theor. Biol.* **455** (2018), pp. 249–260.

[142] J. West *et al.* "Normal tissue architecture determines the evolutionary course of cancer." *Nat. Commun.* **12** (2021).

[143] J. West *et al.* "Towards multidrug adaptive therapy." *Cancer Res.* **80** (2020), pp. 1578–1589.

[144] A. Wu *et al.* "Game theory in the death galaxy: interaction of cancer and stromal cells in tumour microenvironment." *Interface Focus* **4** (2014).

[145] X. S. Xie. "Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics." *J. Chem. Phys.* **117**.24 (2002), pp. 11024–11032.

[146] C. N. Yang and T. D. Lee. "Statistical theory of equations of state and phase transitions. I. Theory of condensation." *Phys. Rev.* **87** (1952), pp. 404–409.

[147] L.-S. Young. "What are SRB measures, and which dynamical systems have them?" *J. Stat. Phys.* **108** (2002), pp. 733–754.

[148] J. Zhou *et al.* "Nonequilibrium population dynamics of phenotype conversion of cancer cells." *PLOS One* **9** (2014).

[149] J. Zhou *et al.* "Quasi-potential landscape in complex multi-stable systems." *J. R. Soc. Interface* **9** (2012), pp. 3539–3553.

[150] P. Zhou and T. Li. "Construction of the landscape for multistable systems: potential landscape, quasi-potential, a-type integral and beyond." *J. Chem. Phys.* **144** (2016).

[151] P. Zhou *et al.* "Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics." *Nat. Commun.* **12** (2021).

[152] R. W. Zwanzig. "High-temperature equation of state by a perturbation method. I. Nonpolar gases." *J. Chem. Phys.* **22**.8 (1954), pp. 1420–1426.

Appendix A

QUASI-POTENTIAL THEORY

A.1 Dominant balance for WKB approximation

In order to determine dominant behavior of solutions f_ε to (2.6), we expand (2.6) using vector calculus identities:

$$\frac{\partial f_\varepsilon}{\partial t} = f_\varepsilon\Big(-\nabla\cdot\mathbf{b}(\mathbf{x})+\varepsilon\sum_{j=1}^{m}\nabla\cdot\frac{\partial D_j}{\partial x_j}\Big)+\nabla f_\varepsilon\cdot\Big(-\mathbf{b}(\mathbf{x})+\varepsilon\sum_{j=1}^{m}\frac{\partial D_j}{\partial x_j}\Big)+\varepsilon\sum_{j=1}^{m}\sum_{i=1}^{m}\frac{\partial f_\varepsilon}{\partial x_j}\frac{\partial D_{ij}}{\partial x_i}+\frac{\partial^2 f_\varepsilon}{\partial x_i x_j}D_{ij}(\mathbf{x}). \tag{A.1}$$

From (2.13), we have:

$$\nabla f_\varepsilon(\mathbf{x},t) = \frac{f_\varepsilon}{\delta}\sum_{n=0}^{\infty}\delta^n\nabla S_n(\mathbf{x},t); \tag{A.2}$$

$$\frac{\partial f_\varepsilon}{\partial x_j} = \frac{f_\varepsilon}{\delta}\sum_{n=0}^{\infty}\delta^n\frac{\partial S_n}{\partial x_j}, \quad \frac{\partial f_\varepsilon}{\partial t} = \frac{f_\varepsilon}{\delta}\sum_{n=0}^{\infty}\delta^n\frac{\partial S_n}{\partial t}; \tag{A.3}$$

$$\frac{\partial^2 f_\varepsilon}{\partial x_i\partial x_j} = \frac{f_\varepsilon}{\delta^2}\Big(\sum_{k=0}^{\infty}\delta^k\frac{\partial S_k}{\partial x_i}\Big)\Big(\sum_{n=0}^{\infty}\delta^n\frac{\partial S_n}{\partial x_j}\Big)+\frac{f_\varepsilon}{\delta}\sum_{n=0}^{\infty}\delta^n\frac{\partial^2 S_n}{\partial x_i\partial x_j}. \tag{A.4}$$

Plugging these expressions into (A.1), we obtain:

$$\frac{f_\varepsilon}{\delta}\sum_{n=0}^{\infty}\delta^n\frac{\partial S_n}{\partial t} = f_\varepsilon\Big(-\nabla\cdot\mathbf{b}(\mathbf{x})+\varepsilon\sum_{j=1}^{m}\nabla\cdot\frac{\partial D_j}{\partial x_j}\Big)+\frac{f_\varepsilon}{\delta}\sum_{n=0}^{\infty}\delta^n\nabla S_n(\mathbf{x},t)\cdot\Big(-\mathbf{b}(\mathbf{x})+\varepsilon\sum_{j=1}^{m}\frac{\partial D_j}{\partial x_j}\Big)$$
$$+\varepsilon\frac{f_\varepsilon}{\delta}\sum_{j=1}^{m}\sum_{i=1}^{m}\left[\frac{\partial D_{ij}}{\partial x_i}\sum_{n=0}^{\infty}\delta^n\frac{\partial S_n}{\partial x_j}+\frac{D_{ij}(\mathbf{x})}{\delta}\left(\frac{\partial S_0}{\partial x_i}\frac{\partial S_0}{\partial x_j}+\delta\frac{\partial S_0}{\partial x_i}\frac{\partial S_1}{\partial x_j}+\delta\frac{\partial S_1}{\partial x_i}\frac{\partial S_0}{\partial x_j}+\cdots+\delta\frac{\partial^2 S_0}{\partial x_j}+\ldots\right)\right] \tag{A.5}$$

Taking this expression to leading order in δ as $\delta \to 0$, and multiplying both sides by δ/f_ε, we get:

$$\frac{\partial S_0}{\partial t} = \delta\Big(-\nabla\cdot\mathbf{b}(\mathbf{x})+\varepsilon\sum_{j=1}^{m}\nabla\cdot\frac{\partial D_j}{\partial x_j}\Big)+\nabla S_0(\mathbf{x},t)\cdot\Big(-\mathbf{b}(\mathbf{x})+\varepsilon\sum_{j=1}^{m}\frac{\partial D_j}{\partial x_j}\Big)+\varepsilon\sum_{j=1}^{m}\sum_{i=1}^{m}\frac{\partial D_{ij}}{\partial x_i}\frac{\partial S_0}{\partial x_j}$$
$$+\varepsilon\sum_{j=1}^{m}\sum_{i=1}^{m}D_{ij}(\mathbf{x})\left[\frac{1}{\delta}\frac{\partial S_0}{\partial x_i}\frac{\partial S_0}{\partial x_j}+\left(\frac{\partial S_0}{\partial x_i}\frac{\partial S_1}{\partial x_j}+\frac{\partial S_1}{\partial x_i}\frac{\partial S_0}{\partial x_j}+\frac{\partial^2 S_0}{\partial x_j}\right)\right] \quad . \quad \text{(A.6)}$$

The dominant balance as $\delta \to 0$ is given by

$$\frac{\partial S_0}{\partial t}+\nabla S_0(\mathbf{x},t)\cdot\mathbf{b}(\mathbf{x}) \sim \frac{\varepsilon}{\delta}\sum_{j=1}^{m}\sum_{i=1}^{m}D_{ij}(\mathbf{x})\frac{\partial S_0}{\partial x_i}\frac{\partial S_0}{\partial x_j} \equiv \frac{\varepsilon}{\delta}\nabla S_0(\mathbf{x},t)\cdot D(\mathbf{x})\nabla S_0(\mathbf{x},t), \quad \text{(A.7)}$$

which tells us that $\delta = \varepsilon$ (i.e., $q = 1$).

A.2 Asymptotic behavior of the Hessian matrix $\mathbf{H}_\varepsilon(\mathbf{x}_0)$

Using the limit definition of $\varphi(\mathbf{x})$, we will show that

$$\ln\det\mathbf{H}_\varepsilon(\mathbf{x}_0) = o(\varepsilon^{-1}) \ ,$$

where $\mathbf{H}_\varepsilon(\mathbf{x}_0)$ denotes the Hessian matrix of $-\varepsilon\ln f_\varepsilon(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}_0$ (which we assume to be positive definite). In particular, we assume that the following limit holds with uniform convergence:

$$\lim_{\varepsilon\to 0} -\varepsilon\ln f_\varepsilon(\mathbf{x}) = \varphi(\mathbf{x}) \ .$$

Because uniform convergence implies pointwise convergence, this limit certainly holds at $\mathbf{x} = \mathbf{x}_0$. Further suppose that the second partial derivatives of $-\ln f_\varepsilon(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}_0$ converge to some limiting value as $\varepsilon \to 0$. Then we can interchange this limit and differentiation to get the following relation:

$$\lim_{\varepsilon\to 0} -\varepsilon\frac{\partial^2}{\partial x_i\partial x_j}\big[\ln f_\varepsilon(\mathbf{x}_0)\big] = \frac{\partial^2}{\partial x_i\partial x_j}\big[\varphi(\mathbf{x}_0)\big] \ ,$$

which tells us that as $\varepsilon \to 0$,

$$\lim_{\varepsilon \to 0} \mathbf{H}_\varepsilon(\mathbf{x}_0) = \mathbf{H}_\varphi(\mathbf{x}_0)$$

Using the fact that we can interchange continuous functions and limits, it follows that

$$\lim_{\varepsilon \to 0} \ln \det \mathbf{H}_\varepsilon(\mathbf{x}_0) = \ln \det \mathbf{H}_\varphi(\mathbf{x}_0) \ .$$

Then we indeed have that $\ln \det \mathbf{H}_\varepsilon(\mathbf{x}_0) = o(\varepsilon^{-1})$, as

$$\lim_{\varepsilon \to 0} \varepsilon \ln \det \mathbf{H}_\varepsilon(\mathbf{x}_0) = \ln \det \mathbf{H}_\varphi(\mathbf{x}_0) \cdot \lim_{\varepsilon \to 0} \varepsilon = 0 \ .$$

Appendix B

CANCER ODE MODEL

B.1 Linear stability analysis

By the general theory of linear ODEs, the exact solution to this system is

$$x(t) = c_1 e^{\lambda_1 t} v^{(1)} + c_2 e^{\lambda_2 t} v^{(2)}\,, \tag{B.1}$$

where $\lambda_i, v^{(i)}$ are the right eigenpairs of A. The eigenvalues of A are given by:

$$\lambda_{1,2} = \frac{g_S + g_R - k_{SR} - k_{RS} \pm \sqrt{(g_S + g_R - k_{SR} - k_{RS})^2 - 4(g_S g_R - g_R k_{SR} - g_S k_{RS})}}{2} \tag{B.2}$$

where we define the net growth rates $g_S = b_S - d_S$and $g_R = b_R - d_R$ for simplicity.

Define $\alpha = g_R - g_S - (k_{RS} - k_{SR})$ and $\beta = \sqrt{\alpha^2 + 4k_{SR}k_{RS}}$. Using the expression given in (B.2), along with the fact that

$$g_S + g_R - k_{SR} - k_{RS} = \alpha + 2(g_S - k_{SR}),$$

we can write λ_1 and λ_2 in terms of α and β:

$$\begin{aligned}
\lambda_{1,2} &= \frac{\alpha + 2(g_S - k_{SR}) \pm \sqrt{(\alpha + 2(g_S - k_{SR})^2 - 4(g_S g_R - g_R k_{SR} - g_S k_{RS})}}{2} \\
&= \frac{\alpha + 2(g_S - k_{SR}) \pm \sqrt{\alpha^2 + 4(\alpha - g_R)(g_S - k_{SR}) + 4(g_S - k_{SR})^2 + 4 g_S k_{RS}}}{2} \\
&= \frac{\alpha + 2(g_S - k_{SR}) \pm \sqrt{\alpha^2 + 4(g_S - k_{SR})(\alpha - g_R + g_S - k_{SR}) + 4 g_S k_{RS}}}{2} \\
&= \frac{\alpha + 2(g_S - k_{SR}) \pm \sqrt{\alpha^2 - 4k_{RS}(g_S - k_{SR}) + 4 g_S k_{RS}}}{2} \\
&= \frac{\alpha + 2(g_S - k_{SR}) \pm \sqrt{\alpha^2 + 4 k_{SR} k_{RS}}}{2} \\
&= \frac{\alpha + 2(g_S - k_{SR}) \pm \beta}{2} .
\end{aligned}$$

Because we assume that k_{SR} and k_{RS} are strictly positive, the discriminant $\alpha^2 + 4k_{SR}k_{RS}$ is strictly positive, and $\beta > 0$. Therefore, both of the above eigenvalues are real and distinct. The corresponding eigenvectors are

$$v^{(1,2)} = \frac{1}{2k_{SR}} \begin{bmatrix} -\alpha \pm \beta \\ 2k_{SR} \end{bmatrix} .$$

The coefficients c_1, c_2 in (B.1) satisfy the linear system $x(0) = Vc$, where V is the matrix whose columns are the eigenvectors of A, and $c = \begin{bmatrix} c_1 & c_2 \end{bmatrix}^T$. It can be shown that

$$c_1 = \frac{k_{SR}}{\beta} x_1(0) + \frac{\alpha + \beta}{2\beta} x_2(0), \quad \text{and} \quad c_2 = -\frac{k_{SR}}{\beta} x_1(0) - \frac{\alpha - \beta}{2\beta} x_2(0) .$$

In the following analysis, we require that $g_R > 0$ under drug treatment as a general definition of the drug-resistant phenotype.

The origin is guaranteed to be a saddle point wherever

$$g := \frac{g_S}{g_R} > \frac{g_S - k_{SR}}{k_{RS}} \tag{B.3}$$

The condition in (B.3) has two interpretations based on whether $g_S > 0$ or $g_S < 0$. Wherever $g_S > 0$, (B.3) requires that at least one of the relative fitness of the drug-sensitive phenotype g and the backflow rate k_{RS} are sufficiently large compared to the net flux $g_S - k_{SR}$. Wherever $g_S < 0$, (B.3) is true when the ratio of the net flux to the backflow rate is larger in magnitude than the relative fitness.

If instead the growth and transition rates satisfy the inequality

$$g = \frac{g_S}{g_R} < \frac{g_S - k_{SR}}{k_{RS}} \tag{B.4}$$

then the relationship between $g_S + g_R$ and $k_{SR} + k_{RS}$ determines whether the origin is stable or unstable (Table B.1). When $g_S + g_R > k_{SR} + k_{RS}$, the origin is an unstable node, and the tumor grows exponentially (Table B.1). On the other hand, when $g_S + g_R < k_{SR} + k_{RS}$, the origin is a stable node, and the tumor tends toward extinction, i.e., treatment has eradicated the tumor (Table B.1).

One way to frame these two cases is to consider which of the two dynamics, growth or phenotype switching, is the dominant force behind the overall population dynamics. Considering the growth of the overall population of cells, we can think of $g_S + g_R$ as the unweighted growth rate of the tumor, as the true growth rate is given by $\frac{dx_1}{dt} + \frac{dx_2}{dt} = g_S x_1 + g_R x_2$. Considering only the pure phenotype switching dynamics, where total size is fixed, the quantity $k_{SR} + k_{RS}$ is the relaxation time of these dynamics. Therefore, when pure growth outweighs pure switching, tumor growth is unchecked, and no tumor control is possible. If, however, the switching dynamics dominate the underlying growth rate, the tumor will eventually be eradicated.

B.2 Dynamics of tumor population recovery

The following analysis follows from the fact that we can express the total population $N(t) = x_1(t) + x_2(t)$ as a linear combination of exponential terms via the exact solution $x(t)$:

$$N(t) = x_1(t) + x_2(t) = c_1\big(v_1^{(1)} + v_2^{(1)}\big)e^{\lambda_1 t} + c_2\big(v_1^{(2)} + v_2^{(2)}\big)e^{\lambda_2 t}\,. \tag{B.5}$$

We can use the exact solution given in (B.5) to derive an approximate expression for t_P, assuming that it is non-zero. If $N'(0) < 0$, $t = t_P$ is the unique non-zero time point for which $N(t) = N(0)$. Using (B.5), the equation $N(t_P) = N(0)$ reads:

$$N(t_P) \equiv c_1\big(v_1^{(1)} + v_2^{(1)}\big)e^{\lambda_1 t_P} + c_2\big(v_1^{(2)} + v_2^{(2)}\big)e^{\lambda_2 t_P} = N(0)\,. \tag{B.6}$$

Solving the above expression analytically for t_P is not possible, as we have established that the eigenvalues $\lambda_{1,2}$ are distinct.

Under the assumption that the exponential term $e^{\lambda_1 t}$ gives the dominant contribution to $N(t)$ at time $t = t_P$, we get:

$$N(t_P) \approx c_1\big(v_1^{(1)} + v_2^{(1)}\big)e^{\lambda_1 t_P} =: N^*\,. \tag{B.7}$$

Setting the above expression equal to $N(0)$ and solving for t_P, we obtain the following approximation t_P^* of t_P :

$$t_P \approx t_P^* := \frac{1}{\lambda_1}\ln\left[\frac{N(0)}{c_1\big(v_1^{(1)} + v_2^{(1)}\big)}\right]\,. \tag{B.8}$$

Taking the absolute value of the difference between $N(t_P)$ and its approximation $N^* := N(t_P^*)$ as a proxy for the absolute error $|t_P - t_P^*|$, we find that the error of the approximation in

(B.7) scales exponentially with the negative eigenvalue λ_2:

$$|N(t_P) - N^*| = \left| c_2\left(v_1^{(2)} + v_2^{(2)}\right) \right| e^{\lambda_2 t_P}. \tag{B.9}$$

In the above derivation, we neglected the negative eigenvalue term to obtain a closed-form approximation for t_P. However, the rate of cancer remission is equally as important in setting the TTP as is the rate of cancer regrowth. Thus, in order to better understand how t_P changes with m, we also consider the point $t = t_{min}$ at which the total population $N(t)$ reaches its global minimum. Unlike with t_P, we can determine a closed form of t_{min} in terms of the model parameters. The turning point of $N(t)$ occurs when

$$0 = \frac{dN}{dt} \equiv g_S x_1(t) + g_R x_2(t) \iff \rho(t) := \frac{x_2(t)}{x_1(t)} = -\frac{g_S}{g_R}\,.$$

Thus, in order to better understand the dynamics of $N(t)$, we must first understand those of the ratio $\rho(t)$ of drug-resistant cells to drug-sensitive cells.

Using the quotient rule, we can compute the derivative of this ratio, $\frac{d\rho}{dt}$:

$$\frac{d\rho}{dt} = \frac{x_1 x_2' - x_2 x_1'}{x_1^2} = \frac{x_2}{x_1}\left(\frac{x_2'}{x_2} - \frac{x_1'}{x_1}\right) = \rho\left(g_R - k_{RS} + \frac{k_{SR}}{\rho} - g_S + k_{SR} - k_{RS}\rho\right)\,.$$

Recall that $\alpha = g_R - g_S - (k_{RS} - k_{SR})$. Then the above expression simplifies to

$$\frac{d\rho}{dt} = -k_{RS}\rho^2 + \alpha\rho + k_{SR}\,. \tag{B.10}$$

This ODE is a Riccati equation with initial condition $\rho_0 := \rho(0) = x_2(0)/x_1(0)$. Define $q(\rho)$ as the polynomial on the right-hand side of (B.10):

$$q(\rho) = -k_{RS}\rho^2 + \alpha\rho + k_{SR}\,.$$

As long as $q(\rho)$ has real roots, we can solve (B.10) analytically via separation of variables.

Recall the definition $\beta = \sqrt{\alpha^2 + 4k_{SR}k_{RS}}$. Then the roots of $q(\rho)$ are given by

$$\rho_{1,2} = \frac{-\alpha \pm \beta}{-2k_{RS}} = \frac{1}{2k_{RS}} (\alpha \mp \beta) \ .$$

We know that $\alpha^2 + 4k_{SR}k_{RS} > 0$, which tells us that the roots $\rho_{1,2}$ are real and distinct. Therefore, we can write $q(\rho)$ in the following factored form:

$$q(\rho) = -k_{RS}(\rho - \rho_1)(\rho - \rho_2).$$

Using the fact that $\rho_1 < \rho_2$, a simple phase-line stability analysis tells us that ρ_1 is an unstable fixed point, and ρ_2 is a stable fixed point.

With this factorization of $q(\rho)$, we can now solve the given IVP to obtain a closed form of $\rho(t)$. We begin by deriving the partial fraction decomposition of $1/q(\rho)$:

$$\begin{aligned}
&\frac{1}{q(\rho)} = \frac{1}{-k_{RS}(\rho - \rho_1)(\rho - \rho_2)} = \frac{C_1}{\rho - \rho_1} + \frac{C_2}{\rho - \rho_2} \\
\iff &1 = -k_{RS}C_1(\rho - \rho_2) - k_{RS}C_2(\rho - \rho_1) \iff -\frac{1}{k_{RS}} = (C_1 + C_2)\rho - C_1\rho_2 - C_2\rho_1 \\
\iff &C_1 + C_2 = 0 \quad \text{and} \quad \frac{1}{k_{RS}} = C_1\rho_2 + C_2\rho_1 \\
\iff &\frac{1}{k_{RS}(\rho_2 - \rho_1)} = C_1 = -C_2.
\end{aligned}$$

The unique solution to the above system of linear equations for C_1 and C_2 is

$$C_1 = -C_2 = \frac{1}{k_{RS}(\rho_2 - \rho_1)} \ .$$

Therefore, the partial fraction decomposition of $1/q(\rho)$ is given by

$$\frac{1}{q(\rho)} = \frac{1}{k_{RS}(\rho_2 - \rho_1)} \left(\frac{1}{\rho - \rho_1} - \frac{1}{\rho - \rho_2} \right) \ .$$

We can now solve the Riccati equation in (B.10) for $\rho(t)$:

$$\frac{d\rho}{dt} = q(\rho) \iff \int \frac{d\rho}{q(\rho)} = \int dt\,,$$

where we use the partial fraction decomposition of $1/q(\rho)$ to evaluate the integral in ρ up to a constant of integration:

$$\int \frac{d\rho}{q(\rho)} = \frac{1}{k_{RS}(\rho_2-\rho_1)} \int \left(\frac{1}{\rho-\rho_1} - \frac{1}{\rho-\rho_2}\right) d\rho = \frac{1}{k_{RS}(\rho_2-\rho_1)} \ln\left(\frac{\rho-\rho_1}{\rho-\rho_2}\right) - C.$$

It follows from this result that

$$\frac{1}{k_{RS}(\rho_2-\rho_1)} \ln\left(\frac{\rho-\rho_1}{\rho-\rho_2}\right) = t + C\,. \tag{B.11}$$

Applying the initial condition $\rho(0) = \rho_0$, we can solve for the constant C:

$$C = \frac{1}{k_{RS}(\rho_2-\rho_1)} \ln\left(\frac{\rho_0-\rho_1}{\rho_0-\rho_2}\right)\,.$$

We can now solve for the exact solution ρ in (B.11):

$$\begin{aligned}
&\ln\left(\frac{\rho-\rho_1}{\rho-\rho_2}\right) = k_{RS}(\rho_2-\rho_1)t + \ln\left(\frac{\rho_0-\rho_1}{\rho_0-\rho_2}\right) \\
\iff &\frac{\rho-\rho_1}{\rho-\rho_2} = \frac{\rho_0-\rho_1}{\rho_0-\rho_2} e^{k_{RS}(\rho_2-\rho_1)t} \\
\iff &\rho\Big(\rho_0-\rho_2-(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t}\Big) = \rho_1(\rho_0-\rho_2) - \rho_2(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t} \\
\iff &\rho(t) = \frac{\rho_1(\rho_0-\rho_2) - \rho_2(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t}}{\rho_0-\rho_2-(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t}}\,.
\end{aligned}$$

From this closed expression, we find an explicit expression for t_{min} by solving the equation

$\rho(t) = -g_S/g_R$:

$$\begin{aligned}
&\rho(t) = \frac{\rho_1(\rho_0-\rho_2) - \rho_2(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t}}{\rho_0-\rho_2-(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t}} = -\frac{g_S}{g_R} \\
\iff & g_R\rho_1(\rho_0-\rho_2) - g_R\rho_2(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t} = -g_S(\rho_0-\rho_2) + g_S(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t} \\
\iff & (g_R\rho_1+g_S)(\rho_0-\rho_2) = (g_R\rho_2+g_S)(\rho_0-\rho_1)e^{k_{RS}(\rho_2-\rho_1)t} \\
\iff & \frac{(g_R\rho_1+g_S)(\rho_0-\rho_2)}{(g_R\rho_2+g_S)(\rho_0-\rho_1)} = e^{k_{RS}(\rho_2-\rho_1)t} \\
\iff & t_{min} := t = \frac{1}{k_{RS}(\rho_2-\rho_1)} \ln\left(\frac{(g_R\rho_1+g_S)(\rho_0-\rho_2)}{(g_R\rho_2+g_S)(\rho_0-\rho_1)}\right)
\end{aligned}$$

We note that in the case where $N(t)$ grows monotonically (i.e., when the origin is an unstable node), the global minimum occurs at $t = 0$, so we take $t_{min} := 0$. If $N(t)$ decays monotonically (i.e., the origin is a stable node), t_{min} is not well defined. For each of Cases A, B, C, and D, we find that as a function of drug dose, t_{min} strongly correlates with t_P (Figures B.1-B.4). We have that $t_P \approx 2t_{min}$ at low drug doses, and $t_P \approx 3t_{min}$ at high drug doses (Figures B.1-B.4).

B.3 Supplemental tables and figures

Table B.1: Summary of all possible cases for the stability of the origin under ODE (3.2).

	$g_S + g_R > k_{SR} + k_{RS}$	$g_S + g_R < k_{SR} + k_{RS}$
$g > \frac{g_S - k_{SR}}{k_{RS}}$	**saddle point**	**saddle point**
$g < \frac{g_S - k_{SR}}{k_{RS}}$	**unstable node**	**stable node**

Table B.2: Initial conditions, rate constants and pharmacodynamic parameters with fixed values (units not specified).

Parameter	Description	Value
$x_1(0)$	concentration of sensitive cells at onset of treatment	200
$x_2(0)$	concentration of resistant cells at onset of treatment	20
b_S	per-capita birth rate of sensitive cells	0.2
b_R	per-capita birth rate of resistant cells	0.1
k_{RS}	"backflow" rate (resistant-to-sensitive)	0.01
$\delta_S = \delta_R$	$\approx$ basal death rate of all cells	0.02
κ	$\approx$ basal transition rate (sensitive-to-resistant)	0.01
$r_S = r_R = r_{SR}$	drug saturation rate	0.2
$P_S = P_R$	EC50 for cell killing	20
E_S	drug efficacy for killing sensitive cells	0.3
E_R	drug efficacy for killing resistant cells	0.04

Table B.3: Parameter distributions used to general virtual patient cohorts (units not specified).

Parameter	Description	Distribution
ξ	fraction of resistant cells at onset of treatment	$\mathcal{U}(0, 1)$
b_S	per-capita birth rate of sensitive cells	$\mathcal{U}(0.1, 0.3)$
b_R	per-capita birth rate of resistant cells	$\mathcal{U}(0.05, 0.15)$
k_{RS}	per-capita "backflow" rate (resistant-to-sensitive)	$\mathcal{U}(0, 0.02)$
$\delta_S = \delta_R$	$\approx$ basal death rate of all cells	$\mathcal{U}(0.01, 0.03)$
κ	$\approx$ basal transition rate (sensitive-to-resistant)	$\mathcal{U}(0, 0.02)$

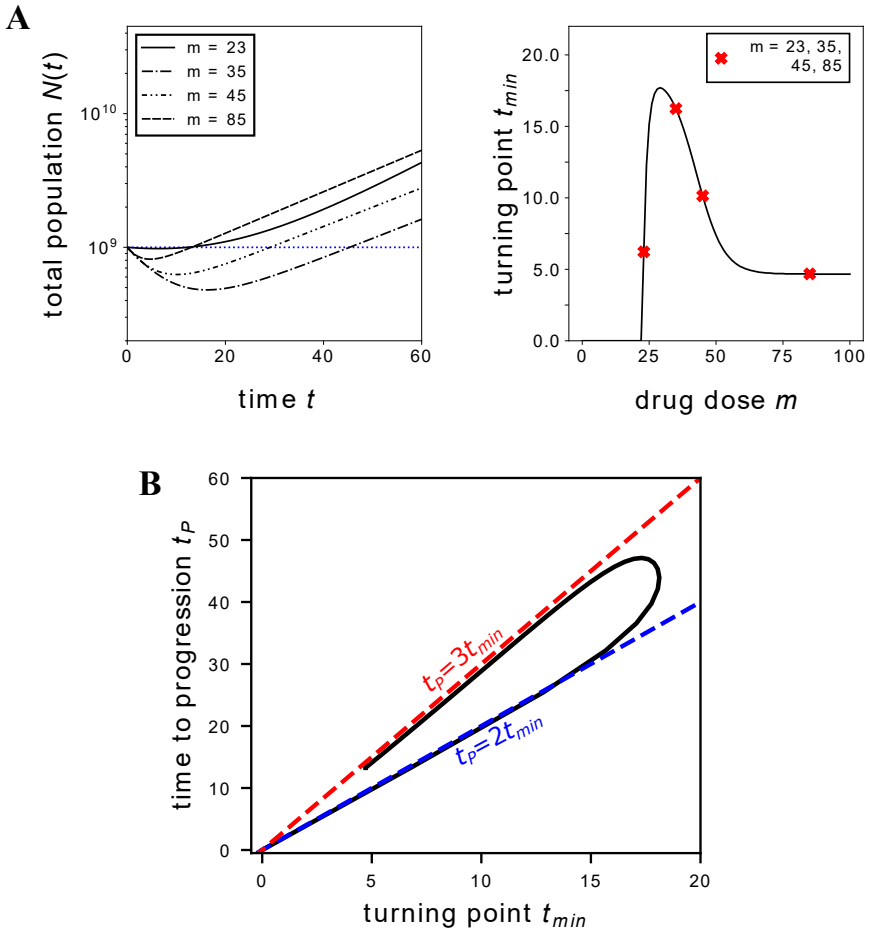

Figure B.1: *Dynamics of tumor population recovery for Case A.* **A.** *Left:* Time-evolution of the total population $N(t)$ plotted on a logarithmic scale for drug doses $m = 23, 35, 45, 85$. The horizontal dashed line indicates initial population size $N(0)$. *Right:* Turning point t_{min}, at which $N(t)$ reaches a minimum, as a function of drug dose m. Red markers indicate reference points of $m = 23, 35, 45, 85$. **B.** Parametric curve $(t_{min}(m), t_P(m))$ relating the turning point t_{min} and the time to progression t_P. Dashed lines of slope 2 (blue) and 3 (red) are given for reference.

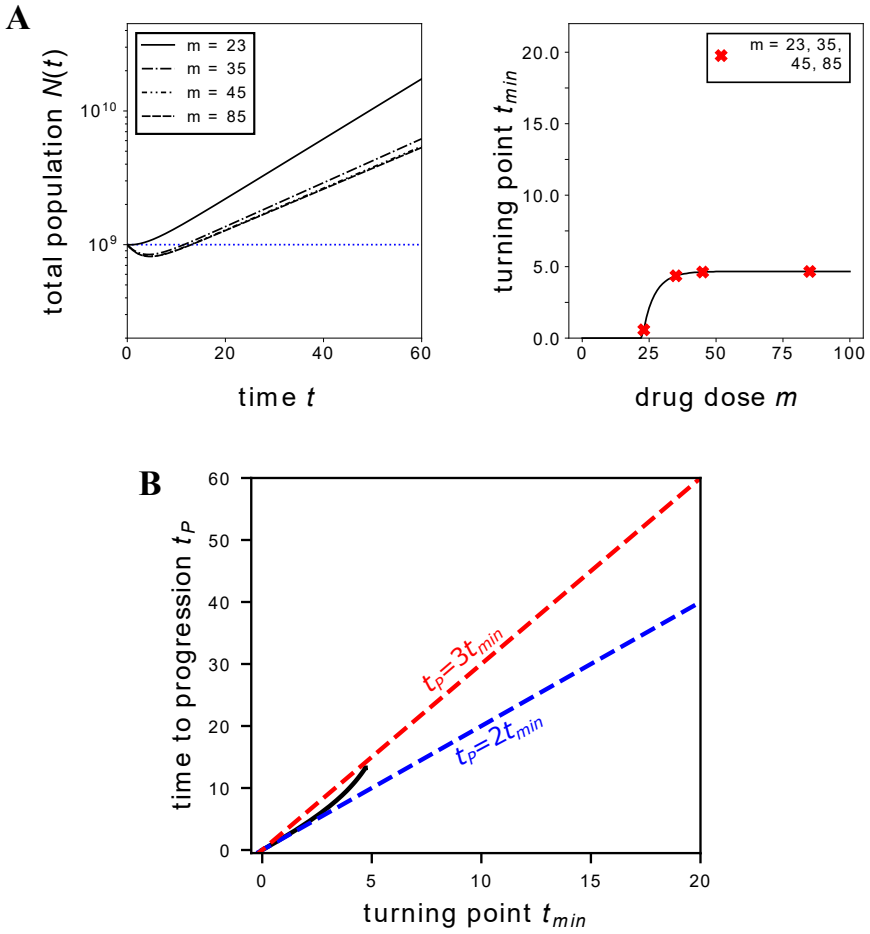

Figure B.2: *Dynamics of tumor population recovery for Case B.* **A.** *Left:* Time-evolution of the total population $N(t)$ plotted on a logarithmic scale for drug doses $m = 23, 35, 45, 85$. The horizontal dashed line indicates initial population size $N(0)$. *Right:* Turning point t_{min}, at which $N(t)$ reaches a minimum, as a function of drug dose m. Red markers indicate reference points of $m = 23, 35, 45, 85$. **B.** Parametric curve $(t_{min}(m), t_P(m))$ relating the turning point t_{min} and the time to progression t_P. Dashed lines of slope 2 (blue) and 3 (red) are given for reference.

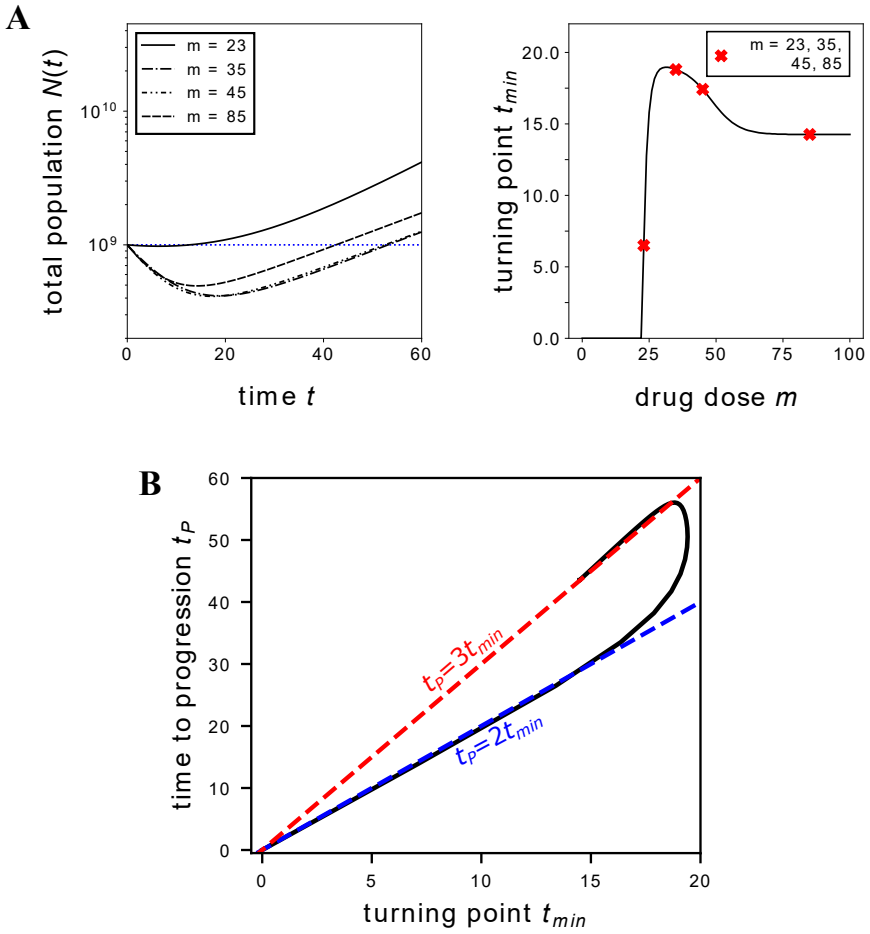

Figure B.3: *Dynamics of tumor population recovery for Case C.* **A.** *Left:* Time-evolution of the total population $N(t)$ plotted on a logarithmic scale for drug doses $m = 23, 35, 45, 85$. The horizontal dashed line indicates initial population size $N(0)$. *Right:* Turning point t_{min}, at which $N(t)$ reaches a minimum, as a function of drug dose m. Red markers indicate reference points of $m = 23, 35, 45, 85$. **B.** Parametric curve $(t_{min}(m), t_P(m))$ relating the turning point t_{min} and the time to progression t_P. Dashed lines of slope 2 (blue) and 3 (red) are given for reference.

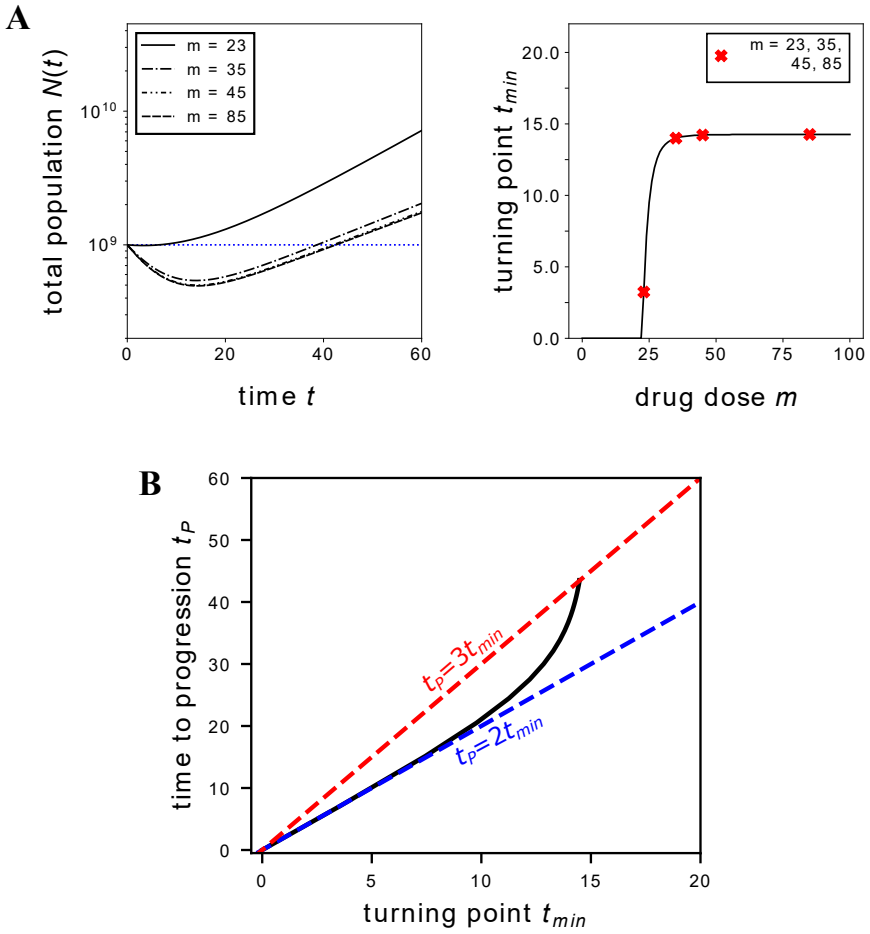

Figure B.4: *Dynamics of tumor population recovery for Case D.* **A.** *Left:* Time-evolution of the total population $N(t)$ plotted on a logarithmic scale for drug doses $m = 23, 35, 45, 85$. The horizontal dashed line indicates initial population size $N(0)$. *Right:* Turning point t_{min}, at which $N(t)$ reaches a minimum, as a function of drug dose m. Red markers indicate reference points of $m = 23, 35, 45, 85$. **B.** Parametric curve $(t_{min}(m), t_P(m))$ relating the turning point t_{min} and the time to progression t_P. Dashed lines of slope 2 (blue) and 3 (red) are given for reference.

Appendix C

STOCHASTIC THERMODYNAMICS

C.1 Tracking vs. counting and entropy of assimilation

We begin by noting that the constraint $m_0+m_1+\cdots+m_n = N$ in the counting representation implies that any m_i is uniquely determined by the other n measurements m_j. Then, taking $\mathbf{x} = (m_0, m_1, \cdots, m_{n-1})$, we can write the multinomial probability $\Pr\big(\{m_i\}\big)$ as

$$\Pr\big(\{m_i\}\big) = \Pr(\mathbf{x}) := C_n(\mathbf{x})\frac{N!}{m_0!\, m_1! \cdots m_{n-1}!}p_0^{m_0}p_1^{m_1}\cdots p_{n-1}^{m_{n-1}} \ ,$$

where

$$C_n(\mathbf{x}) = \frac{q^{M(\mathbf{x})}}{M(\mathbf{x})!}, \quad q = 1 - \sum_{i=0}^{n-1} p_i, \quad M(\mathbf{x}) = N - \sum_{j=1}^{n} x_j \ .$$

Assuming N is sufficiently large so that the Central Limit Theorem holds, we can approximate $\Pr(\mathbf{x})$ with a multinomial Guassian:

$$\Pr(\mathbf{x}) \simeq f(\mathbf{x}) := \frac{1}{Z_N}e^{-\frac{1}{2}(\mathbf{x}-\mu_N)^T\Sigma_N^{-1}(\mathbf{x}-\mu_N)}$$

where

$$Z_N = \sqrt{(2\pi)^n \det(\Sigma_N)}, \quad \mu_N = (Np_0, Np_1, \cdots, Np_{n-1}), \text{ and } \ \big(\Sigma_N\big)_{ij} = Np_i(\delta_{ij} - p_j) \ .$$

Under this assumption, we can approximate the entropy of counting as follows:

$$\begin{aligned} S_{count} &= -\sum_{\{m_i\}} \Pr\big(\{m_i\}\big) \ln \Pr\big(\{m_i\}\big) \\ &\simeq -\int_{\mathbb{R}^n} f(\mathbf{x}) \ln f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &= \ln Z_N + \int_{\mathbb{R}^n} \frac{1}{2} (\mathbf{x} - \mu_N)^T \Sigma_N^{-1} (\mathbf{x} - \mu_N) \, f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \end{aligned}$$

By properties of the determinant, $Z_N \simeq \det(\Sigma_N) = \mathcal{O}(N^n)$, so that $\ln(Z_N) = \mathcal{O}(n \ln N)$. To determine the asymptotic order of the integral term in the above expression, we note that we can re-write the integral as $I(1)$, where

$$I(\xi) := \frac{1}{Z_N} \int_{\mathbb{R}^n} \frac{1}{2} (\mathbf{x} - \mu_N)^T \Sigma_N^{-1} (\mathbf{x} - \mu_N) \, e^{-\frac{\xi}{2} (\mathbf{x} - \mu_N)^T \Sigma_N^{-1} (\mathbf{x} - \mu_N)} \, \mathrm{d}\mathbf{x}$$

We can evaluate the integral $I(\xi)$ as follows:

$$\begin{aligned} I(\xi) &= \frac{1}{Z_N} \int_{\mathbb{R}^n} -\frac{\partial}{\partial \xi} \left[e^{-\frac{\xi}{2} (\mathbf{x} - \mu_N)^T \Sigma_N^{-1} (\mathbf{x} - \mu_N)} \right] \, \mathrm{d}\mathbf{x} \\ &= -\frac{1}{Z_N} \frac{\partial}{\partial \xi} \left[\int_{\mathbb{R}^n} e^{-\frac{\xi}{2} (\mathbf{x} - \mu_N)^T \Sigma_N^{-1} (\mathbf{x} - \mu_N)} \, \mathrm{d}\mathbf{x} \right] \\ &= -\frac{1}{Z_N} \frac{\partial}{\partial \xi} \left[\sqrt{(2\pi\xi)^n \det(\Sigma_N)} \right] \\ &= -\frac{1}{Z_N} \frac{n (2\pi\xi)^{n-1} \det(\Sigma_N)}{2\sqrt{(2\pi\xi)^n \det(\Sigma_N)}} \\ &= -\frac{n (\xi)^{n-1}}{4\pi \sqrt{\xi^n}} \ . \end{aligned}$$

Evaluating this expression at $\xi = 1$, we get that:

$$I(1) = -\frac{n}{4\pi} = \mathcal{O}(n) \ .$$

Therefore, to leading order,

$$S_{count} \simeq \ln Z_N + I(1) = \mathcal{O}(n \ln N).$$

This asymptotic behavior is in sharp contrast with $S_{track} = \mathcal{O}(N) \gg S_{count}$.

C.2 Lagrangian duality: multiplier as a conjugate variable

We start with a representative example to illustrate that the critical point in the method of Lagrange multipliers is a saddle point. Define the bivariate objective function as $f(\mathbf{x}) = a_1 x_1^2 + a_2 x_2^2$, where $a_1, a_2 > 0$, and define the constraint as $g(\mathbf{x}) = b_0 + b_1 x_1 + b_2 x_2 = 0$. Then the Lagrangian function corresponding to minimizing f subject to $g = 0$ is

$$L(x_1, x_2, \alpha) = \big(a_1 x_1^2 + a_2 x_2^2\big) + \alpha\big(b_0 + b_1 x_1 + b_2 x_2\big),$$

where α is the Lagrange multiplier. With fixed constant $x_2 = c$, $L(x_1, c, \alpha) = a_1 x_1^2 + b_1 \alpha x_1 + (b_0 + b_2 c)\alpha + a_2 c^2$ is a saddle for (x_1, α). The same is true for (x_2, α) when we hold x_1 constant, $L(c', x_2, \alpha)$.

Let us now consider the general problem of minimization of a convex objective $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, under the constraint $g(\mathbf{x}) = g^*$:

$$J(g^*) = \min_{\mathbf{x}} \Big\{ f(\mathbf{x}) \,\Big|\, g(\mathbf{x}) = g^* \Big\}.$$

Then the Lagrangian function corresponding to this problem is given by:

$$L(\mathbf{x}, \alpha) = f(\mathbf{x}) + \alpha g(\mathbf{x}),$$

in which the Lagrangian multiplier α is the conjugate variable to g^* via Legendre-Fenchel

transform (LFT). The method of Lagrange multipliers specifies the optimal value $J(g^*)$ via:

$$\begin{cases} H(\alpha) = \min\limits_{\mathbf{x}} L(\mathbf{x}, \alpha) = \min\limits_{\mathbf{x}} \big\{ f(\mathbf{x}) + \alpha g(\mathbf{x}) \big\}, \\ J(g^*) = \max\limits_{\alpha} \big\{ H(\alpha) - \alpha g^* \big\}, \\ H(\alpha) = \min\limits_{g^*} \big\{ J(g^*) + \alpha g^* \big\}, \end{cases}$$

in which $H(\alpha)$ is the (concave) Lagrange dual function and α is related to g^* through

$$\alpha = -\frac{\mathrm{d}J(g^*)}{\mathrm{d}g^*} \;\text{ or }\; g^* = \frac{\mathrm{d}H(\alpha)}{\mathrm{d}\alpha}.$$

Taken together, these equations give us an equation in differential form that represents the solution to the constrained minimization problem:

$$\mathrm{d}_{\mathbf{x}} f(\mathbf{x}^*) + \alpha \, \mathrm{d}_{\mathbf{x}} g(\mathbf{x}^*) = v$$

$$H(\alpha) - J(g^*) - \alpha g^* = 0.$$

C.2.1 Results for Gibbs conditioning

First, we derive the expression given in the main text for the distribution $\nu^*(y; \alpha)$ that minimizes $\Phi[\boldsymbol{\nu}]$ subject to the linear constraint $G[\boldsymbol{\nu}] = \overline{g}$. By Lagrangian duality:

$$\begin{aligned} \phi(\overline{g}) &= \inf_{\boldsymbol{\nu}} \Big\{ \Phi[\boldsymbol{\nu}] \;\Big|\; G[\boldsymbol{\nu}] = \overline{g} \Big\} \\ &= \sup_{\alpha} \inf_{\boldsymbol{\nu}} \Big\{ \Phi[\boldsymbol{\nu}] + \alpha \big(G[\boldsymbol{\nu}] - \overline{g} \big) \Big\} \\ &= \sup_{\alpha} \left\{ -\alpha \overline{g} + \inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] + \alpha \int_{\mathcal{X}} \nu(y) g(y) \mathrm{d}y \right\} \right\} \\ &= \sup_{\alpha} \left\{ -\alpha \overline{g} + \Psi[\alpha \boldsymbol{g}] \right\}. \end{aligned}$$

Therefore, we have the identity

$$\phi(\overline{g}) = -\alpha\overline{g} + \Psi[\alpha\boldsymbol{g}],$$

wherever α satisfies the equality

$$\overline{g} = \sum_{i\in\mathcal{S}} g_i \frac{\partial\Psi}{\partial\mu_i}[\alpha\boldsymbol{g}] = \frac{\displaystyle\int_{\mathcal{X}} g(y)p(y)e^{-\alpha g(y)}\mathrm{d}y}{\displaystyle\int_{\mathcal{X}} p(y)e^{-\alpha g(y)}\mathrm{d}y}. \tag{C.1}$$

Moreover, for

$$\boldsymbol{\nu}(\alpha) := \arg\inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] + \alpha \int_{\mathcal{X}} \nu(y)g(y)\mathrm{d}y \right\}$$

we must have for all $i \in \mathcal{S}$

$$\nabla_{\boldsymbol{\nu}}\Psi\big[\boldsymbol{\nu}(\alpha)\big] = -\alpha\boldsymbol{g}\ ,$$

which gives us the expression

$$\nu(y;\alpha) = \frac{p(y)e^{-\alpha g(y)}}{\displaystyle\int_{\mathcal{X}} p(y)e^{-\alpha g(y)}\mathrm{d}y}\ . \tag{C.2}$$

Taken together, these result tell us that $\boldsymbol{\nu}^*$, defined as

$$\nu^*(y;\alpha) = \arg\inf_{\boldsymbol{\nu}} \left\{ \Phi[\boldsymbol{\nu}] \;\middle|\; G[\boldsymbol{\nu}] = \overline{g} \right\}$$

is given by the expression in (C.2), where α satisfies the condition given in (C.1). We also note that $\boldsymbol{\nu} = \boldsymbol{\nu}^*$ is exactly the density at which $\tilde{\Phi} = 0$ for a given $\overline{g}$.

Now, we will derive the expression for the LFT of $\tilde{\Phi}$, $\tilde{\Psi}$:

$$\begin{aligned}
\tilde{\Psi}\big[\hat{\mathbf{u}}\,|\,\overline{g}\big] &= \inf_{\boldsymbol{\nu}}\left\{\int_{\mathcal{X}}\hat{u}(y)\nu(y)\mathrm{d}y + \tilde{\Phi}\big[\boldsymbol{\nu}\,|\,\overline{g}\big]\right\} \\
&= \inf_{\boldsymbol{\nu}}\left\{\int_{\mathcal{X}}\hat{u}(y)\nu(y)\mathrm{d}y + \Phi[\boldsymbol{\nu}] - \phi(\overline{g}) \;\middle|\; G[\boldsymbol{\nu}] = \overline{g}\right\} \\
&= \inf_{\boldsymbol{\nu}}\left\{\int_{\mathcal{X}}\hat{u}(y)\nu(y)\mathrm{d}y + \Phi[\boldsymbol{\nu}] \;\middle|\; G[\boldsymbol{\nu}] = \overline{g}\right\} - \phi(\overline{g})
\end{aligned}$$

By Lagrangian duality, we have:

$$\begin{aligned}
\inf_{\boldsymbol{\nu}}\left\{\int_{\mathcal{X}}\hat{u}(y)\nu(y)\mathrm{d}y + \Phi[\boldsymbol{\nu}] \;\middle|\; G[\boldsymbol{\nu}] = \overline{g}\right\} &= \inf_{\boldsymbol{\nu}}\left\{\int_{\mathcal{X}}\hat{u}(y)\nu(y)\mathrm{d}y + \delta_0\left(G[\boldsymbol{\nu}] - \overline{g}\right) + \Phi[\boldsymbol{\nu}]\right\} \\
&= \sup_{\alpha}\left\{\inf_{\boldsymbol{\nu}}\left\{\int_{\mathcal{X}}\nu(y)\big(\hat{u}(y) + \alpha g(y)\big)\mathrm{d}y + \Phi[\boldsymbol{\nu}]\right\} - \alpha\overline{g}\right\} \\
&= \sup_{\alpha}\left\{\Psi[\hat{\mathbf{u}} + \alpha\boldsymbol{g}] - \alpha\overline{g}\right\}
\end{aligned}$$

Therefore,

$$\tilde{\Psi}\big[\hat{\mathbf{u}}\,|\,\overline{g}\big] = \Psi[\hat{\mathbf{u}} + \alpha\boldsymbol{g}] - \alpha\overline{g} - \phi(\overline{g}),$$

where α satisfies the condition

$$\overline{g} = \frac{\partial}{\partial\beta}\Big[\Psi\big[\hat{\mathbf{u}} + \beta\boldsymbol{g}\big]\Big] = \frac{\displaystyle\int_{\mathcal{X}} g(y)p(y)e^{-\hat{u}(y)-\alpha g(y)}\mathrm{d}y}{\displaystyle\int_{\mathcal{X}} p(y)e^{-\hat{u}(y)-\alpha g(y)}\mathrm{d}y}\ .$$

These are exactly the expressions given in the main text for $\tilde{\Psi}$.