Chapter 1

Introduction

1.1 Motivation and focus

Biological research, and natural sciences in general, commonly aquire new knowledge by testing a hypothesis against experiment data. A multitude of experimental methods and protocols has been developed over time to reveal the composition of biological systems. Analysing complex biological systems aiming at a deeper understanding of the processes in living systems requires the integration of experimental and computational research (Kitano 2002). Particularly rich information is present in high-dimensional single-cell data. Such data is generated by methods like microscopy, flow cytometry or single-cell RNA sequencing, where the abundance of up to thousands of cellular components for every individual cell in a population is measured. These experiments thus capture the heterogeneity present in a cell population.

When talking about reasons for heterogeneity in biological systems, one often differentiates between intrinsic and extrinsic noise. Intrinsic noise is characterized by the absence of, or only short time correlation between quantities in identical cells of a population. It is thought of as a system inherent property that emerges from stochastic fluctuations in biochemical reactions involving low copy number of genes and thereby causing heterogeneity in a population. Extrinsic noise on the other hand exhibits long time correlations of the quantities in a population (Elowitz et al. 2002; Swain et al. 2002; Munsky et al. 2009; Iversen et al. 2014). Such persistent variance between cells in a population is caused by various factors including the local environment or the history of cells (Snijder et al. 2009; Gut et al. 2015; Sandler et al. 2015). Examples for extrinsic noise caused by history of cells are the cell cycle or cell differentiation processes. Therein, the cellular components vary depending on the progress of individual cells along the cell cycle or differentiation pathway.

Experimental single-cell data that is randomly spread around the population average is observed for a stationary population with prevailing intrinsic noise (Figure 1.1 a). In contrast, single-cell data may be spread around a path in the data space, if cells in the population are additionally at different stages of a process (Figure 1.1 b).

Hence, information about a biological process is present in single-cell data where the population is spread over that process in terms of: (1) shape of the path in data space, (2) the distribution of cells along the path and (3) variance of the population around the path. Furthermore, by changing these characteristics a cell population can be manipulated to achieve a desired behavior. This may for example be inhibition of cell growth in cancer treatment, neuron synchronization in jet-lag, or neuron desynchronization in Parkinson patients.

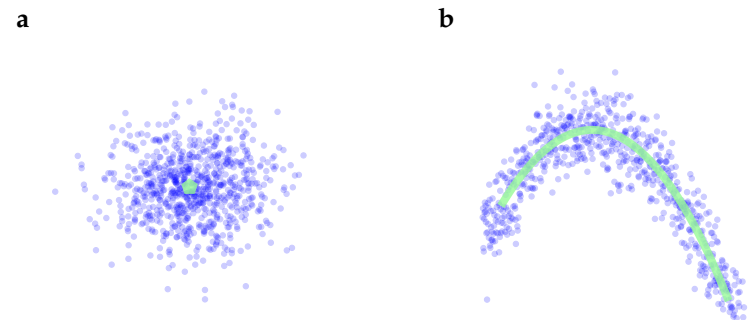

Figure 1.1. Distinction between heterogeneity in single-cell data originating from (**a**) random noise in a stationary cell population or (**b**) an additional underlying process.

Systems and computational biology exploit mathematical models to understand and predict the dynamics of biological systems. A common way to mathematically model molecular processes in a cell is via ordinary differential equations (ODE) models (Klipp et al. 2009; MacArthur et al. 2009). These models describe the concentration change of cellular components in a single cell, or the average cell of a population under deterministic dynamics

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \,, \\ x(0) &= x_0 \,. \end{aligned} \tag{1.1}$$

Therein, the state variables $x(t) \in \mathbb{R}^n$ represent different molecular species in the cell which can be affected by external inputs $u(t) \in \mathbb{R}^l$ such as media, drugs, optogenetic cues or environmental factors. The dynamics are determined by the vector field $f \colon \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n$.

A collection of nearly identical cells, also termed an ensemble, may be modeled as multi-agent system, with each agent being a dynamical system with dynamics given by Eq. (1.1). Mathematically, an ensemble can also be described in terms of a density function over a state space $p(x,t)$ (Wiener 1938) as shown in Figure 1.2. The dynamics are governed by partial differential equations, belonging to the class of *Liouville equations* (Gyllenberg and Webb 1990; Brockett 2012) of the general form

$$\begin{aligned} \partial_t p(x,t) &= -\langle \partial_x,\, f(x,u(t))\, p(x,t) \rangle \,, \\ p(x,0) &= p_0(x) \,, \end{aligned} \tag{1.2}$$

equipped with boundary conditions. The transport equation Eq. (1.2) describes how a density $p_0 \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ of initial states is advected with the flow of a nonlinear differential equation of the form $\dot{x}(t) = f(x(t), u(t))$.

Brockett (2012) defines an ensemble system as a collection of nearly identical dynamical systems which admit a certain degree of heterogeneity, and which are subject to the restriction that they may only be manipulated or observed as a whole. This description is suitable for single-cell experiments where measurement data consists mostly of population snapshots which are sought as representative sample of the population. A population snapshot, taken at a given instance in time, provides a vast number of output measurements. Yet, at the same time information relating a measurement to the individual system that produced the measurement is not provided, since the measurement process usually results in killing the cell, making it impossible to measure that cell again. Furthermore, it is often inherent to the experimental setups or treatment scenarios that cells within the population cannot be

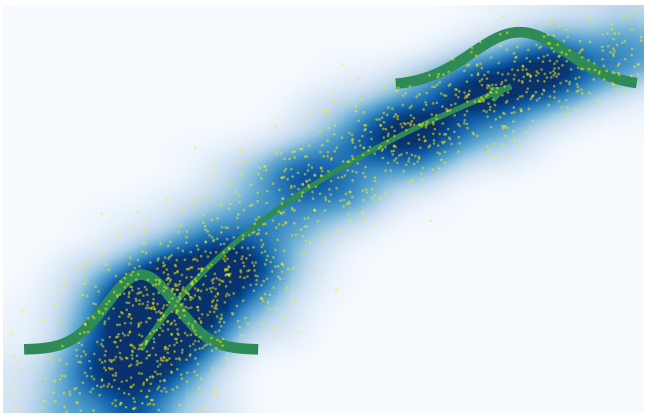

Figure 1.2. Flow of single cells and representation of cellular ensembles.

manipulated individually, but only through a common signal, such as a common stimulus through a drug treatment. Control theory employs mathematical models to derive feedback controllers with the goal to achieve a desired system behavior. The above limitations then lead to the ensemble control problem where we derive a broadcast input signal based on snapshot data to control a heterogeneous cell population.

Major goals of biological and medical research are to (1) understand the dynamics of biological processes, which means to determine the vector field $f(x(t), u(t))$, and (2) control the dynamics, which means to steer the state of a single cell $x(t)$ or a cell population $p(x, t)$ to a desired behavior. This thesis presents analysis and control methods based on single-cell data for cellular processes in heterogeneous populations. We will introduce the underlying concepts in cell cycle studies. In particular, we (1) present a theory to identify the local vector field along a biological process observed in single-cell data and (2) derive an ensemble control formulation to achieve any desired distribution in a population of cellular oscillators.

By reducing the dimensionality of the original data or model to a 1-dimensional manifold, these tasks boil down to problems where we want to analyse or control the distribution of cells along a given process in 1-D. Our results obtained in a 1-dimensional framework are transferable to higher dimensions by a homeomorphism between the description of a process in 1-D and the corresponding high-dimensional model or data.

1.2 Contributions and Outline

In this section we present the outline and summarize the main results and contributions of the individual chapters (illustrated in Figure 1.3). The thesis is structured in two parts: Part I compromising Chapters 2 to 4 contains results on the analysis of biological processes with single-cell data. Part II compromising Chapters 5 and 6 picks up these results for the development of ensemble control algorithms for oscillating cell populations.

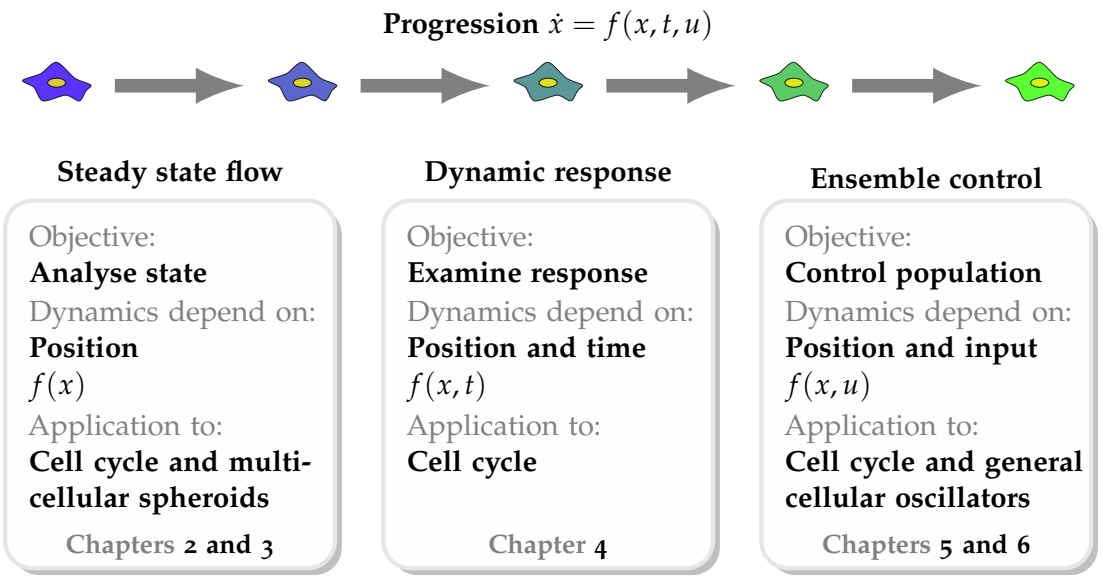

Figure 1.3. Overview of the thesis

Part I Analysis

Chapter 2: Reconstructing temporal and spatial dynamics from single-cell pseudotime

Chapter 2 presents the fundamental theoretic concept to perform real-time analysis with single-cell snapshot data of heterogeneous cell populations spread over different stages of a biological process. First, we introduce the pseudotime representation of single-cell data and briefly discuss its properties and limitations. We then address the arbitrariness of the pseudotime scale by introducing the measure-preserving *map* of pseudotime *i*nto real-*t*ime, in short *MAPiT*. After discussing properties of the method we apply MAPiT for a temporal scale in cell cycle studies and for a spatial scale representing the distance from the surface in cell spheroids. This chapter is based on the publications Kuritz et al. (2017) and Kuritz et al. (2020b). Our main contributions in this chapter are the following:

- MAPiT provides a theoretic basis for the relation of pseudotime values to real temporal and spatial scales.
- MAPiT recovers the progression rate on the process manifold in snapshot data from heterogeneous cell populations.
- By applying MAPiT on two completely distinct problems we demonstrate its universal nature and broad applicability for the analysis of cellular processes.

Chapter 3: Cell cycle analysis with ergodic principles and age-structured population models

Chapter 3 examines the results on cell cycle analysis from Chapter 2 in a dynamical systems perspective. Therein, the evolution of the distribution in pseudotime stems from the description of progression of a single cell through its cell cycle by stochastic differential equation (SDE). Based on ergodic theory, we derive a transformation of this model to age-structured

population models. We do this for scenarios without noise, with extrinsic noise and with intrinsic noise in the description of cell cycle progression. The scenario without noise recapitulates the results from Chapter 2. For the scenarios with noise we derive methods to infer noise strength and incorporate this information in the transformation. Finally, we discuss the different approaches and evaluate the results against live-cell microscopy data. This chapter is based on the publication Kuritz et al. (2017). Our main contributions in this chapter are the following:

- We establish the relation between age-structured population models and cell cycle analysis with snapshot data.
- We derive inference algorithms for intrinsic and extrinsic noise in cell cycle progression.
- We present a transformation from pseudotime to real-time by convolution of the distributions which takes progression noise into account.

Chapter 4: Cell cycle progression inference

Chapter 4 presents an extension of MAPiT to non-stationary processes. The chapter deals with the specific example where we want to infer altered cell cycle progression in response to treatments. We first motivate the problem and describe data processing steps, which are based on MAPiT. Next, we formally describe the partial differential equation (PDE) model and the estimation problem for the inference of a time- and position-dependent cell cycle progression rate. We present a way to efficiently solve the optimization problem by providing parameter sensitivities. Finally we discuss properties of our method and demonstrate its capability with one artificial and two experimental data sets. This chapter is based on the publications Kuritz et al. (2020a). Our main contributions in this chapter are the following:

- We present a computational framework that allows the inference of changes in cell cycle progression from static single-cell measurements.
- We efficient solve the estimation problem for the time- and cell cycle position-dependent progression rate by caclulating parameter sensitivities.

Part II Control

Chapter 5: Passivity-based ensemble control for cell cycle synchronization

Chapter 5 introduces an ensemble control algorithm for cell cycle synchronization. First, we introduce the research topic and formulate the problem in terms of the reduced phase model approach and its relation to MAPiT. Next, we derive the passivity-based control algorithm and provide necessary and sufficient controllability conditions for cell cycle synchronization. Finally, we evaluate the approach in a realistic individual-based simulation framework where we observe parameter ranges in which synchrony is achieved despite the naturally occurring heterogeneity. This chapter is based on the publications Kuritz et al. (2018a) and Kuritz et al. (2018b). Our main contributions in this chapter are the following:

- We introduce a state transformation for age-structured population models to enable passivity-based controller design.

- We derive an ensemble control algorithm to achieve cell cycle synchronization with broadcast input signals.
- We present a theoretic condition for controllability and practical parameters ranges for the synchronization in realistic setups.

Chapter 6: Ensemble control for cellular oscillators

Chapter 6 presents control strategies for the manipulation of processes in heterogeneous cell populations, in particular, cellular oscillators. We introduce a population-level feedback that is capable to achieve any desired distribution of cellular oscillators on their periodic orbit. First, we motivate the research topic including a summary of major results in the field of ensemble control. After deriving the control algorithm we provide controllability conditions which we discuss for some real-world systems. Finally, we present the performance and limitations of the algorithm in computational studies. This chapter is based on the publications Kuritz et al. (2018a) and Kuritz et al. (2019). Our main contributions in this chapter are the following:

- We present an ensemble controller to achieve any distribution of cellular oscillators on their limit cycle.
- We derive controllability conditions for convergence based on properties of the phase response curve.
- Our controller is applicable to many problems, such as, phase shifting of the circadian clock, cell cycle synchronization or desynchronizing of spiking neurons in Parkinson's disease.

Chapter 7: Conclusions

Chapter 7 summarizes the main results of this thesis, presents the conclusions and indicates possible directions for future research.

Appendices

The results in this thesis build up on various theoretic concepts. We describe these concepts in detail in Appendix A. Therein, we briefly introduce the system theoretic basis and control theoretic concepts in Appendix A.1, the concept of reduced phase models in Appendix A.2 and Fourier analysis and circular moments for circular data in Appendix A.3. Furthermore, we cover experimental protocols and data processing procedures in Appendix B, including a review of the basic concepts of trajectory inference algorithms and the resulting pseudotemporal ordering in Appendix B.1. Appendices C and D provide a summary of the spheroid growth model and the cell cycle model, respectively. Finally, Appendix E compromises technical computations and proofs which we do not present in the main chapters in order to improve readability.

The chapters in this thesis are based on severel publications. These publications were addressed to different audiences. For example, Chapter 2 is based on Kuritz et al. (2020b) which had a diverse but application oriented audience in mind. On the other hand, Chapter 6 is based on Kuritz et al. (2019) which is part of a special issue on *Control and Network Theory*

for Biological Systems and thus aimed for a more theory oriented audience. Likewise, we addressed the different chapters in this thesis to different audiences depending on our understanding of the main contributions and their impact on the respective community.

Part I

Analysis

Chapter 2

Reconstructing temporal and spatial dynamics from single-cell pseudotime

This chapter is based on the publication:

> Karsten Kuritz et al. (2020b). 'Reconstructing temporal and spatial dynamics from single-cell pseudotime using prior knowledge of real scale cell densities'. In: *Scientific Reports* 10.1, p. 3619. DOI: 10.1038/s41598-020-60400-z.

Modern cytometry methods allow collecting complex, multi-dimensional data sets from heterogeneous cell populations at single-cell resolution. While methods exist to describe the progression and order of cellular processes from snapshots of such populations, these descriptions are limited to arbitrary pseudotime scales. Deducing real-time dynamics from pseudotemporal ordering however is challenging owing to the arbitrariness of the pseudotime scale. In this chapter, we introduce the measure-preserving *map* of pseudotime *i*nto real-*t*ime, in short *MAPiT*. MAPiT provides a universal transformation method that recovers real-time dynamics of cellular processes from pseudotime scales by utilising knowledge of the distributions on the real scales. As use cases, we applied MAPiT to two prominent problems in the flow-cytometric analysis of heterogeneous cell populations: (1) recovering the spatial arrangement of cells within multi-cellular spheroids prior to spheroid dissociation for cytometric analysis, and (2) recovering the kinetics of cell cycle progression in unsynchronised and thus unperturbed cell populations. Multicellular spheroids grown from cancer cells are widely used as avascular tumour models and proved to be a valuable experimental system, closing the gap between *in vitro* and *in vivo* studies. However, cumbersome preparation of spheroid slices for imaging experiments with limited availability of fluorescent probes restricts the practicability of spheroid experiments. By recovering the spatial position MAPiT reverts the loss of spatial information in single-cell experiments. This enables high-throughput and high-content studies of 3-D-spheroid models. Since MAPiT provides a theoretic basis for the relation of pseudotime values to real temporal and spatial scales, it can be used broadly in the analysis of cellular processes with snapshot data from heterogeneous cell populations.

The experimental data that we present in this chapter was prepared by the Morrison Lab at the Institute of Cell Biology and Immunology at the University of Stuttgart. This chapter is taken in parts from Kuritz et al. (2020b).

2.1 Background and problem formulation

Here, we briefly introduce the pseudotime representation of single cell data and state its main shortcoming which will lead to the problem formulation in this chapter. We provide a comprehensive discussion of the concept of trajectory inference algorithms in Appendix B.1. Single-cell experiments such as flow cytometry, mass cytometry and single-cell RNA-sequencing (scRNA-seq) capture the heterogeneity in cell populations (Klein et al. 2015; Bandura et al. 2009) . The heterogeneity may originate from the fact that the measured cell population is distributed across intermediate cellular states of a biological process, such as cell cycle or cell differentiation (Saelens et al. 2019). This enables the study of biological processes with pseudotime algorithms like CALISTA (Papili Gao et al. 2019), Wanderlust (Bendall et al. 2014), Monocle (Trapnell et al. 2014) or diffusion maps (Haghverdi et al. 2014). These algorithms can capture trajectories in data space by recovering a low-dimensional structure in high-dimensional observations. By ordering cells on a lower dimensional process manifold in the dataspace pseudotime algorithms provide access to the sequence of steps during the process. Common pseudotime algorithms order cells on a pseudotime scale based on a distance metric in the data space, and this metric differs between algorithms (Saelens et al. 2019). Pseudotime values furthermore strongly depend on the measured cellular components. The derived pseudotime thus is a quantitative value of the progression through a biological processes. It is characterized by the relations of high-dimensional observations and in general not equal to the true (time) scale (Weinreb et al. 2018). Hence pseudotime does not directly correspond to real time but is rather a metric in data space of measured cell states. This leads us to the following problem formulation:

Problem 2.1. Given a single-cell snapshot data with a pseudotemporal ordering from a biological process, find a mapping from pseudotime scale to the true (time) scale.

To solve this problem and overcome the arbitrariness of pseudotime scales, we developed *MAPiT* (measure-preserving *map* of pseudotime *i*nto real-*t*ime). MAPiT makes use of prior knowledge of the distribution of cells on the real scale to derive the requested transformation (Figure 2.1). We demonstrate MAPiT for a temporal scale in cell cycle studies and for a spatial scale representing the distance from the surface in multi-cellular tumour spheroids (MCTS). MAPiT robustly reconstructs the true scale of both processes which we verified with imaging data.

2.2 MAPiT - *MAP* of pseudotime *i*nto real-*T*ime

This section presents the theoretic concept for MAPiT and discusses some practical implications.

2.2.1 Measure-preserving transformation for probability distributions

The theoretic foundation of MAPiT originates from measure and probability theory. MAPiT is based on a "measure-preserving transformation" which ensures that the area under the curve is conserved when transforming a probability distribution. Consider a measure space $(X, \mathcal{L}, \lambda)$, where X is a set, $\mathcal{L}$ is a $\sigma-$ ring of measurable subsets of X, and λ is the

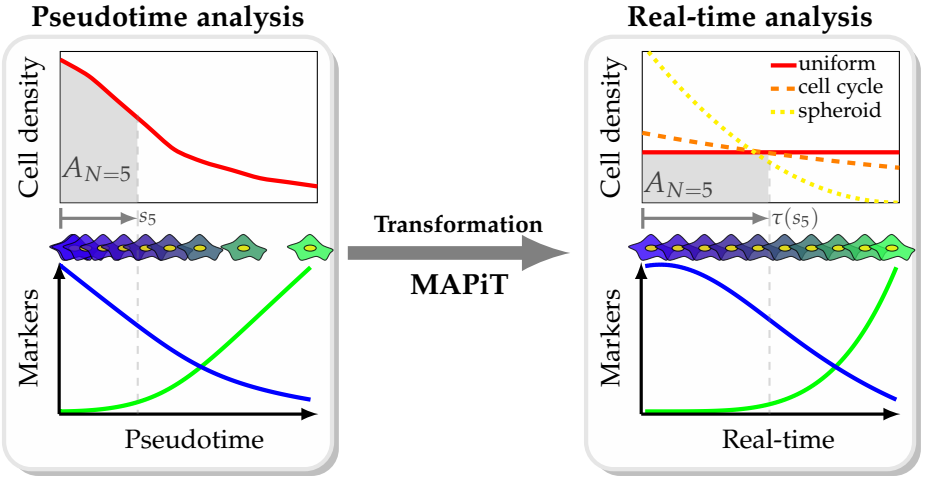

Figure 2.1. MAPiT deduces process dynamics from single-cell snapshot data. Cell density and marker trajectories on pseudotime scale vary with the distance measure used by the pseudotime algorithm and real temporal trajectories cannot be deduced. Cell density, order and trajectories for two markers on pseudotime scale are shown for an exemplary process. As an example pseudotime position of the fifth displayed cell s_5 and associated area under the cell density curve $A_{N=5}$ are indicated in gray. Nonlinear transformation of pseudotime scale recovers true scale dynamics. MAPiT uses prior knowledge of cell densities on the real scale to transform pseudotime to real time by enforcing equality for the area under the density curves at corresponding points on both scales (gray areas). Cell order and marker trajectories are shown for an exemplary uniform distribution on the real scale. Positions of cells across the cell cycle (dashed, orange) or decreasing number of cells towards the center of spheroid cultures (dotted, yellow) are other real scale densities.