Analysis, calibration and evaluation of stochastic models of gene expression

Elias Ventre

Numéro National de Thèse : 2022ENSL0018

THESE

en vue de l'obtention du grade de Docteur, délivré par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N°512
Ecole Doctorale en Informatique et Mathématiques de Lyon

Discipline : Mathématiques

Soutenue publiquement le 28/09/2022, par :
Elias VENTRE

Analyse, calibration et évaluation de modèles stochastiques d'expression des gènes

---

Devant le jury composé de :

LECLERCQ SAMSON, Adeline	PR	Université Grenoble Alpes	Rapporteure
MALRIEU, Florent	PR	Institut Denis Poisson	Rapporteur
YVINEC, Romain	CR HDR	INRAE Tours	Examinateur
ROPERS, Delphine	CR HDR	INRIA Grenoble	Examinatrice
GENTIL, Ivan	PR	Institut Camille Jordan	Examinateur
GANDRILLON, Olivier	DR	ENS Lyon	Directeur de thèse
LEPOUTRE, Thomas	CR HDR	INRIA Lyon	Co-directeur de thèse
ESPINASSE, Thibault	MCF	Institut Camille Jordan	Co-encadrant

Remerciements

Un grand merci à tous ceux qui m'ont aidé dans cet assez long parcours de réorientation qu'a été ma reprise d'étude et cette thèse, car ces 5 années ont vraiment été à la hauteur de mes attentes: stimulantes, agréables, pleines de bonnes recontres !

Tout d'abord, merci à **Eloi** et à **Pascale** qui, bien que non scientifiques, ont donné le coup de pouce me remettant sur la voie des sciences, d'abord par simple plaisir, puis plus sérieusement (et sans perdre le plaisir). Merci à **Nicolas** et **Frédéric**, mes collègues dans le bâtiment, qui ont fait que ces années en entreprise soient enrichissantes et m'ont donné confiance pour ma reconversion. A **Jöel** pour ses conseils et encouragements au moment de reprendre les maths.
Une fois arrivé en master, j'ai eu la chance d'avoir comme enseignant **Philippe Caldero**, qui en plus d'être un enseignant passionant, a su m' orienter vers les bonnes personnes pour me faire découvrir la recherche appliquée. Merci à **Laurent Pujo-Menjouet** qui a pris le temps de m'écouter et de me mettre en contact avec le petit monde des maths-bio à Lyon. A partir de là, j'ai vite recontré ceux qui allaient devenir mes encadrants, d'abord de stage puis de thèse. Je vous suis extrêmement reconnaissant pour avoir fait de ces 3 années (et quelques) de thèse une période passionante, stimulante, et agréable ! Pour la confiance accordée, et les conseils, aides, explications, orientations. **Thibault Espinasse**, pour ta vivacité qui permet de transformer n'importe quelle question en un passionant sujet de recherche, et m'avoir suivi dans les nombreux domaines où j'ai pu m'aventurer, parfois très loin de ta zone de confort (si ce terme peut exister pour toi !). **Olivier Gandrillon**, pour ta pédagogie à l'égard des mathématiciens et ton intuition hors-norme qui amènent à se questionner sérieusement sur la réalité des vies antérieures. Et enfin **Thomas Lepoutre**, pour tes bons conseils scientifiques, professionels ou relationnels, et pour agir toujours dans l'intérêt de tes étudiants et collègues. Un grand merci aussi à ceux avec qui j'ai eu le chance de travailler et de collaborer. **Ulysse Herbach**, car déjà sans ton travail ma thèse n'aurait tout simplement pas existé; **Charles-Edouard Bréhier** et **Vincent Calvez** pour votre grande aide sur mon premier projet, qui est toujours le plus difficile; et particulièrement **Aymeric Baradat** pour tout ce que tu m'as appris au cours de cette troisième année. J'espère vraiment qu'on pourra continuer à travailler ensemble pendant le postdoc !
Je n'oublie évidemment pas **Camille**, **Matteo** et tous les jeunes du LBMC, pour le soutien, les discussions scientifiques ou politiques stimulantes et les bières au foyer, **Maxime** pour l'élevage de blobs, **Philippe, Gérard et Olivia** pour les jeux de rôles et **Geneviève** pour les discussions passionantes et les ballades à la campagne, ainsi que ceux de l'ICJ et en particulier **Simon** pour la préparation des séances d'enseignement. A ce propos, j'ai eu la chance de donner des cours dans les UE dirigées par **Guillaume Aubrun**, **Laurent Bétermin**, **Theresia Eisenkölbl** et **Vincent Borrelli**: vos cours de grande qualité m'ont vraiment aidé pour les ACE !
Je remercie bien sûr **Romain Yvinec** et **Ivan Gentil**, pour avoir fait partie de mon comité de thèse ainsi que de mon jury, ainsi que **Delphine Ropers**, et surtout **Adeline Leclercq-Samson** et **Florent Malrieu** pour avoir accepté de rapporter ma thèse !

Et pour finir, **Pascale** et **Jean-Claude**, **Raphaël** et **Victoire**, pour tout le soutien, les bons moments en confinement, ainsi que la maison de Samoëns, ses vaches et ses moutons.

Contents

Introduction

The difference between cells does not come from the DNA that it contains, since all of an organism's cells contain the same DNA, but from the genes that it expresses. Within a eukaryotic cell's nucleus, DNA is first transcribed into an intermediary form, called mRNA (messenger-RNA). The transcription occurs only when an enzyme called RNA polymerase can bind to a region of the DNA which is called promoter. mRNA molecules are then transported out of the nucleus, where they are translated into proteins (see Figure 1). Proteins will then carry out cell functions like carrying oxygen through the blood, giving structure to tissue, or recognizing specific antigens. Observing these functional properties allows to assign each cell a certain phenotype, called a cell type. We call differentiation the process whereby a cell acquires a specific phenotype, by differential gene expression over time. Considering the huge amount of reactions that take place within a cell and determine its differentiation, a fundamental and complex question in cell biology is then to understand **the mechanisms by which cells differentiate to a cell type or another**, and, as a corollary, how they are stabilized or destabilized.

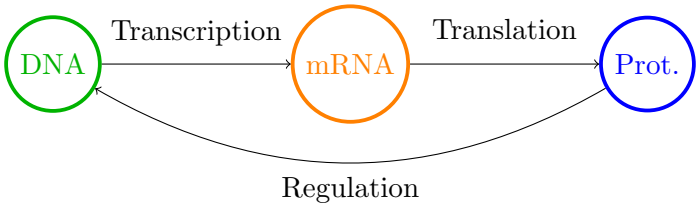


Figure 1: Gene expression mechanism within a single cell.

In addition to carrying out vital functions, gene expression products (namely, proteins) also interact both directly and indirectly with the DNA, and therefore with the expression mechanisms (as represented by the regulation arrow in Figure 1). In particular, the expression of one gene can lead to the activation or inhibition of another. These interactions are commonly referred to as gene regulation and are known to play a central role in the development and differentiation of cells.
The underlying topology of these interactions can be represented by a Gene Regulatory Network (GRN), a powerful concept for representing the interactions between genes through their protein levels. A classical GRN, which is known to play an important role in many biological processes, is the toggle-switch: 2 genes that active themselves and inhibit each other. it can be represented by a graph θ, the weight associated to each of its edges representing the nature of the interactions, as illustrated in Figure 2. The previous questions can then be made more precise: **in a given organism, what GRN drive the differentiation of cells ?** Note however that what we call GRN in this context is an abstraction, that does not only take into account direct interactions, as the binding of a protein on a promoter, but also indirect effects as well as many factors, such as proliferation or cell-to-cell communication that could influence GRN's action.

These issues are difficult to address, and this is mainly due to the nature of the data that have been and are now available. Indeed, the most accessible data are mRNA levels, also

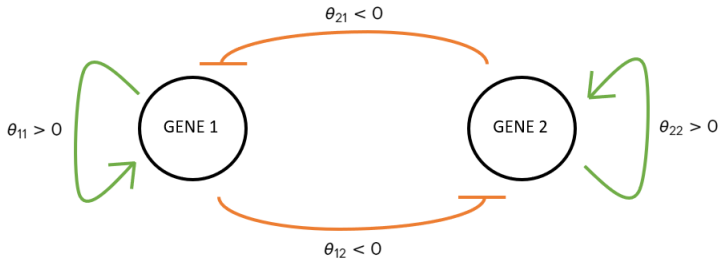


Figure 2: Example of a graph represented a toggle-switch GRN between two genes that activate themselves and inhibit each other, which is represented by the weight matrix $\theta \in \mathbb{R}^{2\times 2}$.

called gene expression measurements or transcriptomic data. They have long been limited to population-based measurements, that is the observation of the mean of gene expression products among a (potentially high) number of cells. Until the 2000's and beyond, most biologists were thinking differentiation as a nearly deterministic process: the genes expressed by a cell and their intensity were thought to be deterministically determined by its DNA and its environment. When it became possible to observe mRNA levels within single cells, about twenty years ago, biologists established that there is a high cell-to-cell variability in gene expression [77], even between cells with the same DNA and in the same environment. This variability has also been shown to be not Gaussian [58]: it seems thus that it is not only due to the accumulation of a large amount of small factor making the cells to slightly deviate from their mean behavior, and that it is biologically relevant when examining a differentiation process. Interestingly, it is now widely accepted that this non-Gaussian variability is mainly due to the bursty nature of mRNAs synthesis [71], also known as *transcriptional bursting* phenomenon, making trajectories of mRNAs of an individual cell very far from those of a diffusion process and particularly difficult to reconstruct from partial observations of the system. Thus, differentiation can be considered as a highly stochastic process [43]: a group of cells with the same DNA and in a same environment are not likely to behave in a similar way. Thus, the mechanisms underlying cell differentiation cannot be understood by the observation of their mean behavior [46] and single-cell data has been available for a relatively short time, which explains that many questions are still to address. Furthermore, even single-cell data does not allow an easy understanding of cell differentiation mechanisms. Indeed, stochasticity prevents the attribution of precise causes to an observed dynamics: a precise analysis of the latter can only be done by using adapted statistical tools, and possible conclusions are meaningful within a certain confidence interval. Moreover, performing such dynamical analysis would require the real trajectories of cell expression products during differentiation. Importantly, this is not possible since the observation process necessarily kills the cells of interest: we have only access to independent samples of cells collected at various time-points instead of real trajectories. We will call these type of data *time-stamped datasets* in the rest of the manuscript.
In that context, we can reformulate once again our previous fundamental biological issue in a more practical way, asking **how, from time-stamped datasets, we can reconstruct the GRN driving the differentiation of the cells that are observed**. An additional question would be also to ask what part of the process is it able to explain.

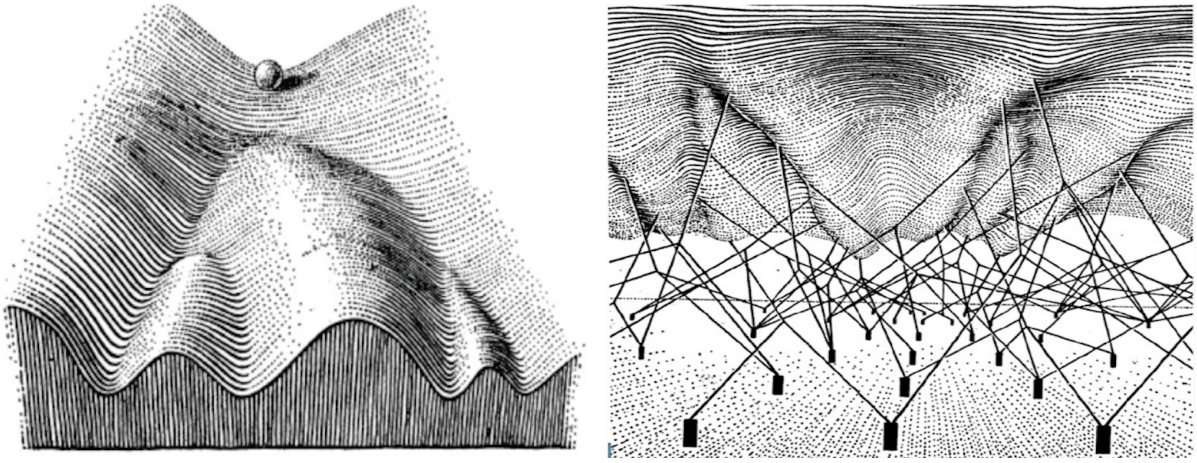
From biology to mathematics

Since measurements technologies now allow the expression of thousands of genes to be measured at the same time, bringing "big data" to biology, it is clear that statistical methods are naturally suited for analyzes. There exists a huge range of statistical methods for analyzing these data in various directions, and it would be difficult to enumerate them. We can mention methods for

representing data in a lower-dimensional space (like the most classical PCA, or recent non-linear methods like UMAP [62]), for analyzing the effect of conditions (analysis of variance, mixture effect models), for measuring correlations between genes or cells (using Pearson or Spearman coefficient or copula), or for ordering the observations along a pseudo-temporal axis [89].

Two types of methods would be of particular interest for us in the following. On one side, to get information about the mechanisms that drive the dynamics of an individual cell, a solution would be to reconstruct most-probable trajectories of differentiation from one or many independent samples of cells collected at various time-points along a developmental progression. Many methods have been developed for reconstructing such trajectories [93, 104], allowing to identify key mechanisms driving certain biological processes [84]. On the other side, under the previous hypothesis that the dynamics of cells mainly results from the action of an underlying GRN, another way of getting information about the underlying mechanisms would to reconstruct this GRN from experimental datasets. To this aim, a large amount of bioinformatics tools have been developed those last twenty years [41, 3], both for population or single-cell data. However, all these methods are almost systematically "top-down" (they do not take into account explicitly the molecular complexity within single-cells) instead of "bottom-up" (based on the molecular processes driving cells dynamics): although they allow to fill some gaps in the information provided by the single-cell data, it is tendentious to interpret their result in a mechanistic way. They also have a major drawback, that is their incompatibility. Indeed, these two categories of methods have actually the same aim, which is the understanding of gene expression patterns observed in an experimental dataset as the emergent property of an underlying GRN. However, the probabilistic models used for trajectories reconstruction are generally too simple to explain their dynamics by the action of a GRN, while most GRN inference methods does not allow to build a reliable link between the inferred network and cell dynamics.
In that context, an important challenge is to develop a bottom-up approach of cell differentiation, able to link the molecular mechanisms within a cell, whose effects are parameterised by the GRN structure, to the dynamics of this cell. It should also allow to analyze time-stamped datasets.

In our work, we are often going to refer to a popular metaphor for understanding cell dynamics, which is the notion of *Waddington's landscape*, introduced by Waddington himself in 1942 [100]. At that time where the physical nature of genes was not yet well known, Waddington developed the metaphor that cells could be seen as marbles following probabilistic trajectories, as they roll through a developmental landscape of ridges and valleys (see Figure 3A) for the original drawing). These trajectories takes place in a so-called gene expression space, describing the space of possible mRNAs and/or proteins levels for all of the genes that are expressed (or observed) within a cell. Following our previous considerations, the resulting landscape can be considered to be mainly shaped by an underlying GRN: interestingly it was already the vision of Washington, who represented the interactions between genes (even if it was imagined before this notion was well defined) giving to the landscape its structure (see Figure 3B). Moreover, the variability observed at the individual cell level argues for a probabilistic description of cell differentiation processes. In that context, the developmental landscape introduced by Waddington can be defined as "a time-varying distribution on gene expression space" [84]: these distributions characterize areas of low and high probabilities in the gene expression space, determining the probable fate of cells in this landscape. Relating this distribution to a GRN should allow to build a mathematical framework allowing to link the mechanisms that drive differentiation processes to the estimation of these time-varying distributions, from time-stamped datasets.



(a) **(b)**

Figure 3: Waddington epigenetic landscape: reproduction from [100].

From the molecular to the functional level

The variability observed in gene expression has also led to question the traditional notion of cell types. Indeed, if the behavior of a cell is subject to stochasticity, how to explain the fact that cells can be distinguished *a posteriori* by their functional properties ? Moreover, although this concept has to be interrogated in the era of single-cell omics [18], cell types seem to remain most of the time stable over time in an organism. More precisely, differentiation is now considered as a metastable process, which is evidenced by the limited number of existing cellular phenotypes [67, 11] and the possibility for a cell to trans-differentiate from on type to another, these transition being rare events. This question sheds light on the link existing between cell biology and statistical physics, yet mentioned in [46]. Indeed, it is a commonplace for physicians that random phenomena at a molecular level can lead to macroscopic structures (which can be the cell types at the single-cell level or the formation of organoids or other macro-organization at a broader scale): there is no paradox between the stochasticity of gene expression on the scale of a cell and the fact that the development of an organism is structured and even reproducible. In this point of view, we can consider that the behavior of a cell is an emergent property of the underlying GRN driving its differentiation. The cell types can be seen as macrostates associated to an underlying mechanistic model describing the behavior of a cell at the molecular level, driven by the GRN. This paradigm has been developed in the case of a diffusion process in [39, 108], where the authors define the cell types as the stable basins of attraction associated to the deterministic drift of a stochastic differential equation modeling gene expression dynamics. We can thus modify our fundamental question once again: **Is it possible to understand, from a GRN, the diversity of cell types observed in an organism as an emergent property the molecular processes induced by its action, and how to reconstruct it from experimental observations ?**

Context and challenges

We can now detail the context of our PhD. In this work, we are going to consider a cell as a complex system, the functional properties of which emerge from the complex molecular processes acting on it, represented by the action of a GRN. Starting from a mechanistic model describing these complex multivariate processes, we are going to embrace a statistical physics point of view on cell differentiation in order to understand the resulting functional behavior of a cell in probabilistic terms. We should also be able to consider the reverse problem of reconstructing a GRN from observable transcriptomic profiles, in such a way that the associated functional

behavior is consistent with the observations. Although this approach is not completely new -analogies between biology and statistical physics for understanding cell differentiation processes have been often proposed those last few years [102, 108, 13, 84]- it has been realized nearly exclusively by considering that differentiation can be well described by the most common stochastic processes used in this field, that is stochastic differential equations (SDEs). For such processes, existing mathematical tools indeed allow to understand the emergent properties of stochastic process (using Large deviations theory [92]), or for reconstructing the landscape of a given process from experimental observations (using Optimal transport theory [75]). But these processes are known to be not in adequacy with the highly stochastic and bursty nature of observable transcriptomic profiles, and it is often not clear how the mathematical theories coming from statistical physics can be extended to realistic models of cellular differentiation. Thus, a mechanistic extension of these theories able to capture the nature of biological processes is still to develop [91].

Overview of the manuscript

In Chapter 1, we are going to present the theoretical context of the questions developed in this introduction, as well as the mathematical notions that we are going to use throughout the thesis. We will also detail some state-of-the-art methods and approaches developed those last few years that have been important references for our personal achievements, either because we used them as a starting point for our work, or because we tried to overcome their limits. In particular, we are going to detail the main probabilistic model of gene expression dynamics, that has been developed by Ulysse Herbach during its PhD within the same team, and that we will use as a building brick throughout the manuscript. In Chapter 2, we will develop an analytical link between the GRN dynamics at the molecular level associated to this model, and the resulting functional behavior of an individual cell. In Chapter 3, we will use these results for developing a numerical method able to reconstruct a GRN from gene expression data, that we will apply on both *in silico* generated and experimental dataset from the literature in Chapter 4. At this stage, we would have then addressed most of the questions stated in this introduction, that is the link between molecular and functional scales of cellular dynamics, and how to characterize this dynamics as the action of an underlying GRN using single-cell datasets. The last two chapters consist in the preliminary development of a mathematical method able to evaluate the accuracy of the model with respect to experimental data, while remaining at a level of precision consistent with the nature of the available data.

We present in Figure 4 a graphical abstract of these PhD projects, with the main mathematical field that will be used. In the following chapter, we are going to introduce these mathematical notions and detail the technical framework and existing methods which will serve as a starting point for our projects.
We present in Figures 5-6-7 an overview of the different projects of this manuscript, illustrating each part with a graphical abstract. Once again, the mathematical background for understanding the content of each picture will be exposed in Chapter 1. We also hope that it may be useful to return to these figures at the end of each section for a synthesis.

Regarding our contributions, Chapter 1 does not contain any original work, and Chapters 2, 3, 4 and 5 contain articles that have been published or submitted for publication. They are presented in their publication format, at the cost of sometimes redundant definitions or presentations of models. Chapter 6 contains the preliminary results of an ongoing project, which has been initiated with Aymeric Baradat.

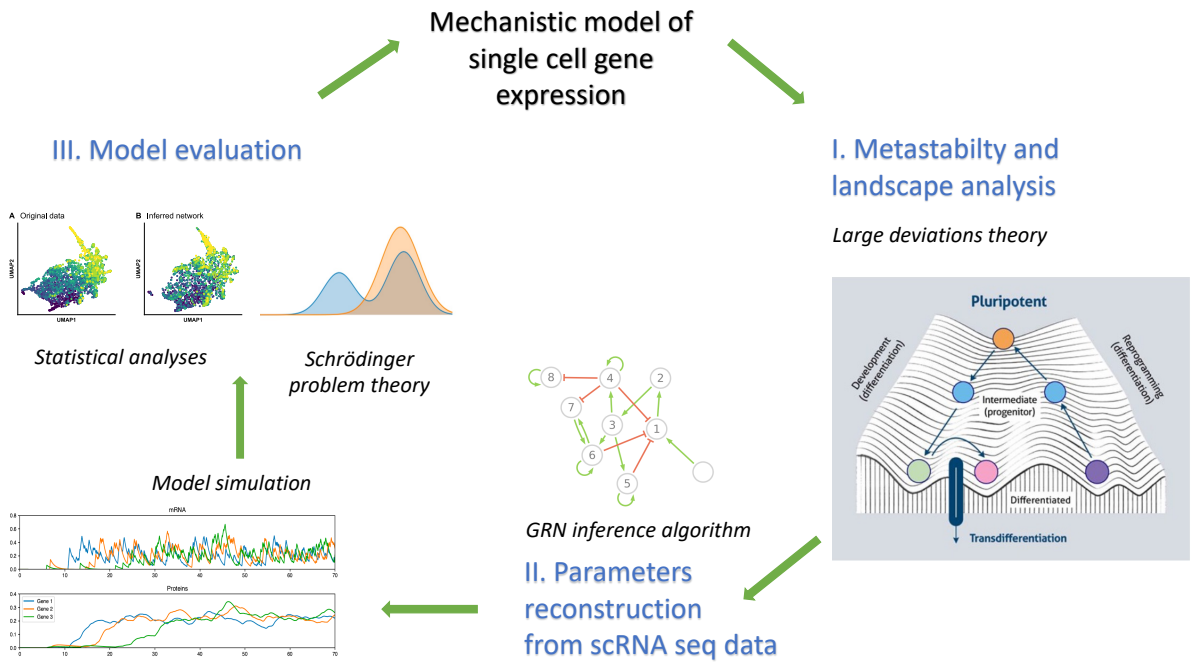


Figure 4: Global view of the PhD. In Part I, starting from a mechanistic model of gene expression, we will use Large deviations theory for reducing this model, in order to link a GRN with an explicit approximate description of the associated "Waddington" landscape. This reduction will be used in Part II to conceive a reverse-engineering method, allowing in particular to infer a most-likely GRN from time-stamped datasets. Combined to some information available in the literature, this method also allow to simulate realistic data that mimicks experimental ones. This allow in particular to evaluate the model against experimental datasets, that we do first using standard statistical tools like Wasserstein distances and dimension reduction techniques, and then the Schrödinger problem in Part III.

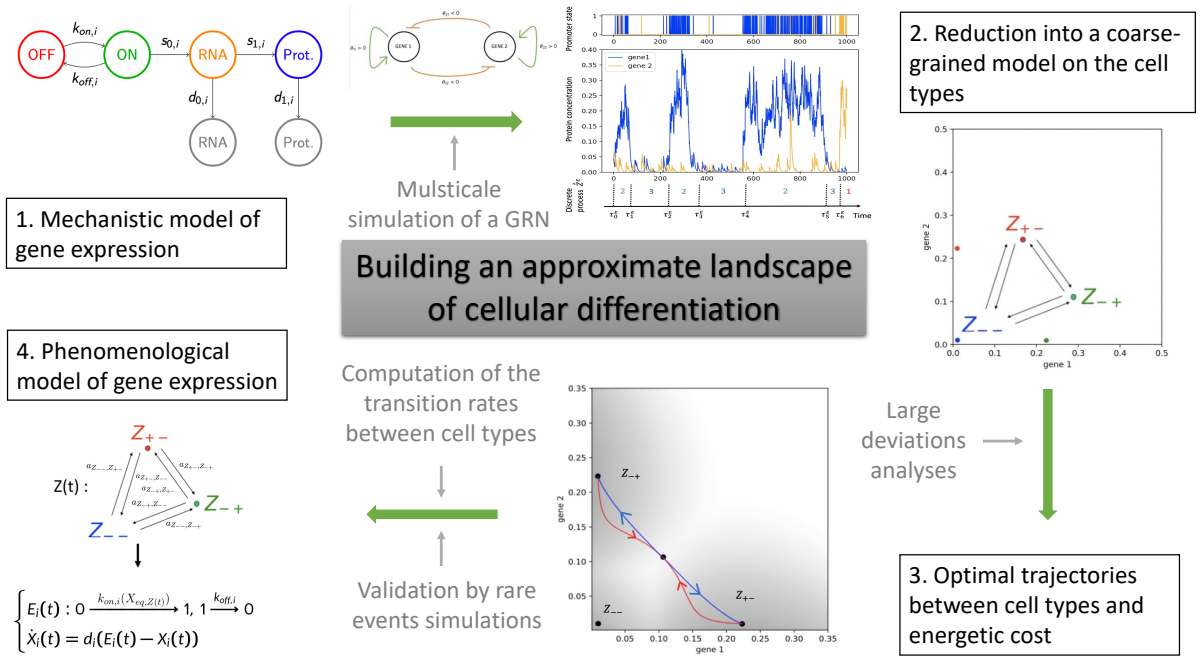


Figure 5: Global view of Part 1. In Chapter 2, we are going to perform a Large deviations analysis of this model in order to reduce it, from any GRN, to a discrete Markov chain on the basins of attraction seen as cell type. We approximate its jump rates by integrating the Lagrangian on the optimal trajectories which characterize the transitions between the attractors of the basins. We then deduce a phenomenological model describing protein dynamics, the stationary distribution of which is explicitly known.

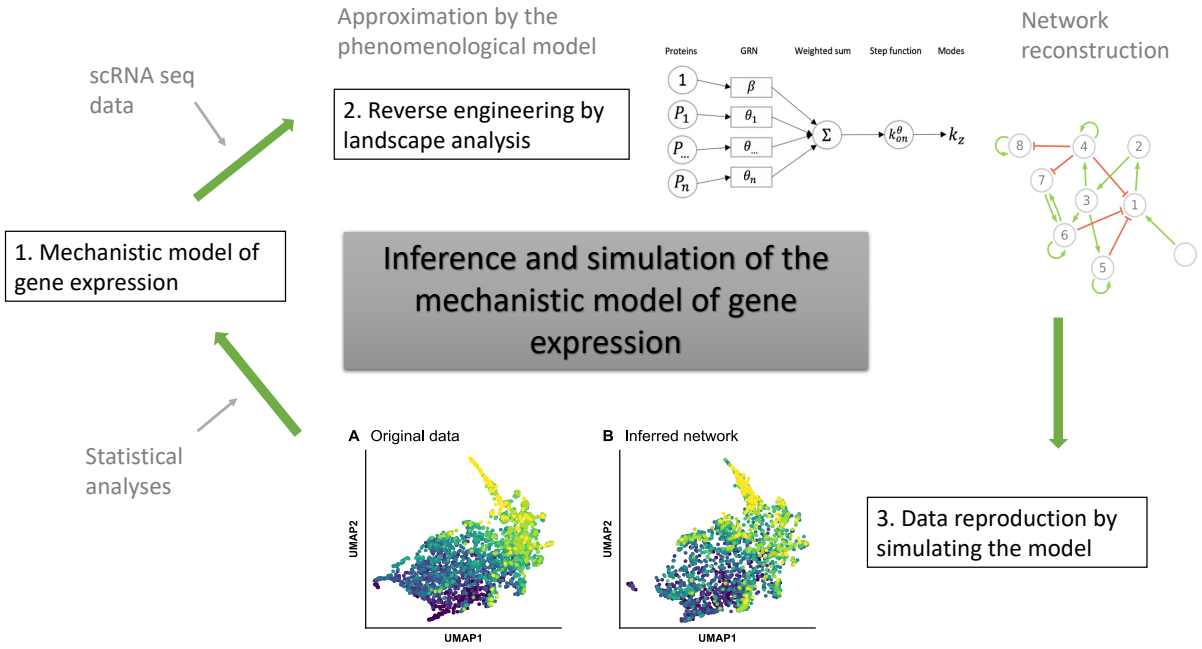


Figure 6: Global view of Part 2. The phenomenological model developed in Part I will be used to conceive in Chapter 3 an original reverse-engineering method from time-stamped scRNA-seq datasets. The method transforms the inference on a set of regression problems, which appears each as the learning step of a simple perceptron. In Chapter 4, we demonstrate the efficiency of this method on both *in silico* generated and experimental datasets, by comparing the results to other state-of-the-art methods and analyzing their biological interpretability.

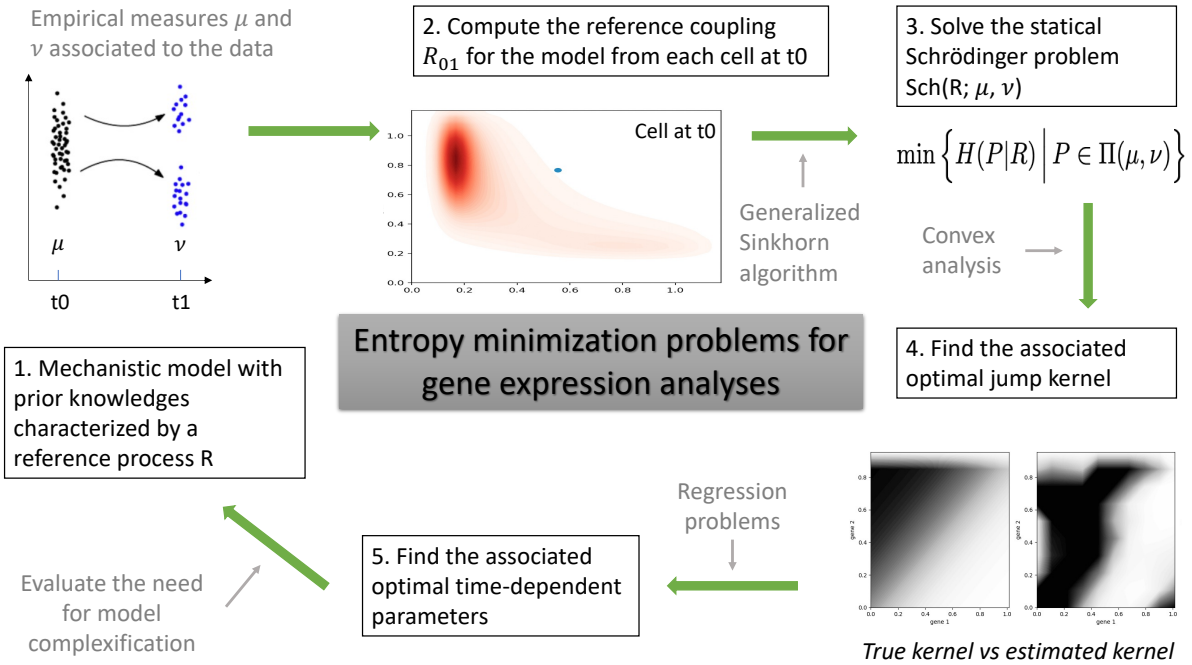


Figure 7: Global view of Part 3. Motivated by the need of a new metric for characterizing the distance between the mechanistic stochastic process and experimental observations, we develop the theoretical basis for solving entropy minimization problems (called Schrödinger problems) when the mechanistic model is used as a reference. We use this method for assessing the quality of the model with respect to experimental observations and a calibrated reference process.

Chapter 1

Mathematical preliminaries and existing related methods.

In this first chapter, we introduce the principal definitions and mathematical notions that are going to be used throughout the manuscript, as well as some state-of-the-art methods that are relevant to contextualize the rest of our work. The aim is to present the works pre-existing to our thesis, which initiated or guided the orientation of our research, and whose limits we tried to identify and overcome.

1.1 A probabilistic point of view on cellular differentiation

1.1.1 Landscape of cellular differentiation

All along this manuscript, we will be interested in the dynamical behavior of a cell undergoing a differentiation process. We first need to describe the space in which we study this dynamics. We consider that a cell evolving in the so-called *gene expression space* is represented by a vector $X = (X_1, \cdots, X_n)$ where each X_i represents the expression level of the i^{th} gene, and n is the number of genes of interest. For the sake of simplicity, we consider that every gene is associated to an unique type of mRNA and protein. Remark however that this is generally not the case, mainly due to phenomena such as alternative splicing: it is considered for example that in mammalian tissues, there are at least 4 times more mRNA types and proteins than genes. However, although important for understanding cell dynamics [64], theses different versions of mRNAs and proteins associated to a gene can be generally identified from experimental data in such a way that we can estimate the quantity of mRNAs and proteins associated to each gene: the study of such phenomena is thus beyond the scope of our studies. In that context, every X_i is then itself a vector of size two: $X_i = (M_i, P_i)$, where M_i and P_i are the levels of mRNA and protein associated to gene i that are in the organism. We call a trajectory $(X_t)_{t\in[0,T]}$ the trajectory of a cell in the gene expression space between times 0 and T.

This formalism motivated the development of dynamical models based on differential equations for describing cell trajectories in the gene expression space, the flow of which depends on the dynamical parameters involved in the chemical reactions driving differentiation [65]. Interestingly, this dynamical formulation allowed to give a first mathematical description of the Waddington's landscape landscape. Every position taken by the cell (seen as a marble) in the gene expression space represents a cellular state, i.e a vector X characterizing the level of expressed mRNAs and proteins in our case. Although the number of possible states is infinite, the number of phenotype observed are limited: for the human, there are about 200 cell types, defined generally by its functional properties. In the representation of Figure 3A, the final basins of low elevation where a high proportion of cells end up correspond to these experimentally observable, terminal cell

types. In that point of view, a marble starts its journey in a valley at the back of the landscape and as it progresses in the gene expression space, it might face branching points along the path, representing the series of fate choices made by a developing cell, before reaching a terminal cell state. Thus, the formalism of dynamical systems gives a clear interpretation of this notion of landscape : if a cell X evolves according to a differential system of equation

$$\dot{X}(t) = F(X(t)),$$

the trajectories would be the one generated by this equation, depending only on the initial conditions and the cell types to the basin associated to each stable equilibrium of the system. The branching points would corresponds to the unstable equilibria, at the boundary of the different basin of attraction, and the slope would represent the velocity of the process along time. Moreover, if there exists a function $V : E \to \mathbb{R}$ such that $F = -\nabla V$, the landscape is shaped by this potential function.

Following the probabilistic point of view developed in the introduction, such deterministic system can only describe average trajectories in the gene expression space, and not individual cell trajectories. When studying a probabilistic system, this is not the evolution of the state which is deterministic, but the evolution of its probability. To give a mathematical formalism to the notion of epigenetic landscape which takes into account stochasticity, we should not consider that there are equations describing trajectories of individual cells in the gene expression space, but instead equations describing the trajectories of cells distribution in the space of probability measures that take values in the gene expression space. In that context, the elevation of the epigenetic surface (the surface representing the landscape in the sens described before) should reflect the probability of observing a particular state in the gene expression space: states that have the highest probability locally will have lower potential and hence will act as the valley-bottoms on the landscape, surrounded by a basin of attraction which would correspond to cells with slightly different states but exhibiting the same phenotype [66]. However, the notion of probability is not clear at this stage: indeed, the probability of a system evolves over time, and so should the landscape, making Waddington's picture less obvious. As we will see later, the solution is often to consider that the landscape is characterised by the steady-state probability of the system, and that this probability also characterizes some of its dynamical aspects.

The most popular approach to describe the epigenetic landscape where cells are subject to stochasticity is to model cell differentiation process by a system of stochastic differential equations (SDEs):

$$\mathrm{d}X_t = F(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t. \tag{1.1}$$

In analogy with the deterministic case, if there exists a potential function V such that $F = -\nabla V$, this function and the diffusion coefficient σ characterize the epigenetic landscape [108]. If not, several methods and approximations have been proposed to find potential functions, generally decomposing the drift in a potential and a rotational part which would characterizes the flux [101, 13], that we will detail in Section 1.3.

In parallel, for a general stochastic system, one common definition of landscape is related to the steady-state distribution $\hat{u}$ (whenever there exists) through the relation $V(x) = -\ln\left(\hat{u}(x)\right)$ [19]. For the so-called Smoluchowski SDE $\mathrm{d}X_t = -\nabla V(X_t) + \sqrt{2\varepsilon}\mathrm{d}B_t$, both definitions coincide exactly since in this case the stationary distribution is $\hat{u}(x) \propto e^{-\frac{V(x)}{\varepsilon}}$.
This function thus characterizes well the areas of the gene expression space with the lowest potential values as the most probable states of a cell at the equilibrium, and its minimum as low probable states, the difference of potential between these two types of areas characterizing an energetic barrier between them [109].

However, it is no longer the case for more complex stochastic systems, for which a general notion of potential function (and landscape) linking the dynamical and the stationary points of view is still to define.

As mentioned in the introduction, the landscape is often regarded to be shaped by an underlying GRN, which can be itself influenced by many factors like cell-to-cell communication, proliferation, etc. In the framework of SDEs, a relevant way to link the dynamics of differentiation processes with a GRN would therefore be to build a model allowing to parameterize the potential function by the latter. However, the nature of the chemical processes driving cell differentiation makes SDE-based model inaccurate for describing the dynamics of cells in the gene expression space. In particular, we have mentioned in the introduction that the bursty synthesis of mRNAs gives rise to highly variable and non-Gaussian expression profiles [58], that would be more in adequacy with the class of *switching ODEs* [9] than with diffusion processes [4, 38]. For this class of stochastic processes, the characterization of a landscape is also more complicated, and requires to adapt the large panel of mathematical tools that are used in statistical mechanics, with the aim of understanding the macroscopic behavior of a system involving a large amount of microscopic sub-processes. Note now that the first goal of this thesis, which will be achieved in Chapter 2 and serve as a starting point for the algorithmic methods developed afterwards, will precisely be this landscape characterization for a specific mechanistic model of gene expression.

1.1.2 Mathematical preliminaries

In this section, we will introduce the principal mathematical notions that we will use throughout the manuscript. They are mainly related to the field of *Stochastic calculus*, but also to the fields of *Schrödinger problem* and *Optimal transport*, whose connections have been intensively studied the last few years. The aim of this section is to provide a simple (and sometimes not completely rigorous) presentation of these mathematical theories, as well as some insights on the way they are going to be interesting for addressing the questions that drive the PhD.

Measures

For introducing Probability theory, we are going to use the notion of measures. If we denote Ω a measurable topological space, we denote $\mathcal{P}(\Omega)$ the set of all its probability measures, *i.e* the positive measures that sums to 1 on Ω. We will use in many situations the notion of pushforward measure:

Definition 1. Let $(\Omega; m)$ a measure space (m is a measure) and E a measurable space. Let us consider $f : \Omega \to E$ a measurable application from Ω to E. The pushforward measure of m by f is denoted $f_{\#}m$. It is a measure on E defined for every subset $A \subset E$ by:

$$f_{\#}m(A) = m(f^{-1}(A)),$$

with $f^{-1}(A) := \{\omega \in \Omega,\, f(\omega) \in A\}$.

For every measurable and positive application ϕ on E, we have:

$$\int_E \phi(y)\mathrm{d}f_{\#}m(\mathrm{d}y) = \int_\Omega \phi \circ f(x)\mathrm{d}m(\mathrm{d}x).$$

Observe that if m is a probability measure and f is positive on E, $f_{\#}m$ is the law of f, that we can see as a random variable with value in E.

Finally, when we consider a stochastic process taking values in E (as we will detail in the following section), we generally identify the random variable describing the process at time t to

the time projection of the measure $P \in C([0,T];E)$ at this time, denoting $X_{t\#}P$ the measure defined by:

$$\forall t > 0,\ X_{t\#}P(\cdot) := P(X_t \in \cdot) \in \mathcal{P}(E).$$

Some notions about stochastic processes

We are interested in the theory of stochastic processes. In a general setting, considering $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space ($\mathcal{F}$ denotes the σ-algebra of Ω), $(E, \mathcal{E})$ a measurable space ($\mathcal{E}$ denotes the σ-algebra of E), and a real number $T > 0$, the application $X : [0,T] \times \Omega \to E$ is said to be a stochastic process defined on Ω, indexed by $[0,T]$ and with values in E if for all $t \in [0,T]$, the application $\Omega \to X(t,\Omega)$ is measurable from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$. We denote by $(X_t)_{t\in[0,T]}$ the random variable characterizing such stochastic process in $[0,T]$, and it will also sometimes denote a realization (when there is no confusion).

Let us denote $(\mathcal{F})_{t\in[0,T]}$ a filtration in $\mathcal{F}$. We define a Markov stochastic process by a transition semi-group $(Q_t)_{t\in[0,T]}$ such that for all t: $Q_t : E \times \mathcal{E} \to [0,1]$ satisfying classical semi-group properties:

- For all $x \in E$, the application $A \to Q_t(x,A)$ defines a probability on $(E, \mathcal{E})$;
- For all $A \in E$: $(t,x) \to Q_t(x,A)$ is measurable for $\mathcal{B}(\mathbb{R}^+) \times \mathcal{E}$;
- For all $x \in E$: $Q_0(x,dy) = \delta_x(y)$;
- For all $s,t > 0$ and $A \in E$: $Q_{t+s}(x,A) = \int_E Q_s(y,A)Q_t(x,dy)$.

A Markov stochastic process relatively to $(\mathcal{F}_t)_{t\in[0,T]}$ with semi-group $(Q_t)_{t\in[0,T]}$ is a $(\mathcal{F}_t)_{t\in[0,T]}$-adapted process on F with values in E such that for all $s,t > 0$ and for all function $f : E \to \mathbb{R}$ measurable and bounded:

$$\mathbb{E}\left(X_{t+s}|\mathcal{F}_s\right) = \int_E f(y)Q_t(X_s, \mathrm{d}y) := Q_t f(X_s),$$

where $\mathbb{E}$ is the expected value under the probability measure $\mathbb{P}$.

When E is a polish space and $(Q_t)_{t\in[0,T]}$ satisfies some regularity properties (for all $f \in C_0(E)$, $Q_t f \in C_0(E)$ and $||Q_t f - f|| \to 0$ as $t \to 0$), we say that $(Q_t)_{t\in[0,T]}$ is a Feller semi-group, in which case the stochastic process $(X_t)_t$ can be characterized by its generator L defined by:

$$\mathcal{L}f := \lim_{t\to 0^+} \frac{Q_t f - f}{t},$$

for all $f \in \mathcal{D}(\mathcal{L})$, $\mathcal{D}(\mathcal{L})$ being the set of functions in $C_0(E)$ such that the limit is well defined on $C_0(E)$. We call a Markov process associated to a Feller semi-group a Feller Markov process. In this manuscript, all the random variables will be defined implicitly on the same probability space with values on a metric space E (which will be often $\mathbb{R}^n$). As they will have a natural construction, the Markov processes that we will consider will be systematically associated to a Feller semi-group, and be then characterized by their generator $\mathcal{L}$.

Finally, Markov Feller processes are often characterized by martingale properties. We recall that a Markov process $(X_t)_{t\in[0,T]}$ relatively to $(\mathcal{F}_t)_{t\in[0,T]}$ called a martingale (w.r.t the filtration $(\mathcal{F}_t)_{t\in[0,T]}$) if $(X_t)_{t\in[0,T]}$ is integrable and for all $0 < s < t$:

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s.$$

It is well known that with the previous notation, if $(X_t)_{t\in[0,T]}$ is a Markov Feller process, then for every $f \in \mathcal{D}(\mathcal{L})$ the stochastic process $(M_{f_t})_{t\in[0,T]}$ defined for all $t \geq 0$ by

$$M_{f_t} = f(X_t) - f(X_0) - \int_0^t Lf(X_s)ds, \tag{1.2}$$

is a martingale w.r.t $(\mathcal{F}_t)_{t\in[0,T]}$ under $\mathbb{P}$. This is relatively simple to see using the semi-group properties of Q_t and the Hille-Yosida theorem which states that for all t:

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_t f = \mathcal{L}Q_t f.$$

Conversely, we say that $(X_t)_{t\in[0,T]}$ is the solution of the martingal problem associated to the generator $\mathcal{L}$ if for every $f \in \mathcal{D}(\mathcal{L})$, the process defined by (1.2) is a is a martingale w.r.t $(\mathcal{F}_t)_{t\in[0,T]}$.

The probability distribution of a Markov Feller process is the solution of the so-called master equation of the process, defined for all $t > 0, x \in E$ by:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t,x) = \mathcal{L}^*\rho(t,x),$$

$\mathcal{L}^*$ being the adjoint operator of $\mathcal{L}$ in $\mathcal{D}(L)$. In plain word, the state of a Markov Feller process is random, but the evolution of its probability distribution is deterministic: knowing an initial distribution $\rho(0,\cdot)$, there is an unique time-dependent distribution ρ which is the solution of the master equation associated to the process.

As in this thesis, we want to deal with experimental observations, we are never going to know the *real* probability distributions of a stochastic process, which could only be characterized by an infinite number of observations. We will nevertheless approximate them from the observations, and try to deduce information about the forces (i.e the GRN in our case) driving the stochastic processes that are observed. A powerful theory for studying the gap between an expected process and its observation, and its consequences, is the theory of Large deviations that we are going to present now.

A first Large deviations principle and its link with the relative entropy

In this section, we introduce the notion of Large deviations when the number of observations of a system tends to infinity, and its link with the relative entropy justified by the Sanov theorem.

When we observe a serie of N independent and identically distributed random variables $(X^n)_{n\leq N}$, following a law $R \in \mathcal{P}(E)$, we expect from the central limit theorem that, when N is high, $(X^n)_{n\leq N}$ follows a Gaussian distribution centered on the mean of the law R, the variance of which is in $O(\frac{1}{N})$. In particular, the empirical distribution

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N}\delta_{X^n},$$

verifies:

$$\int_E x\mu_N(dx) = \frac{1}{N}\sum_{n=1}^{N}X^n \underset{N\to\infty}{\to} \mathbb{E}_R(X).$$

A rare event, when N is large, could be then that $(X^n)_{n\leq N}$ are such that $|\frac{1}{N}\sum_{n=1}^{N}X^n - E_R(X)| > \varepsilon$, with $\varepsilon > 0$.

Equivalently, if we observe a large number N of independent realizations of a stochastic process $((X_t)^n_{t\in[0,T]})_{n\in\mathbb{N}}$ starting from a same initial probability distribution ρ_0, we expect at any time T the collection of values $((X_T)^n)_{n\leq N}$ to follow a Gaussian distribution centered on $\mathbb{E}(X_t|X_0 \sim \rho_0)$ with a variance in $O(\frac{1}{N})$, and a rare event could be that the mean trajectory observed is at a strictly positive distance from the trajectory whose value at each time is the previous expected value.

When we face a rare event in practice, it is natural to question if the observations are very unlikely or the reference measure has to be modified. A question that arises is thus the following: given the observations, what can we say about the reference measure R under which the data were supposed to be chosen? This issue has been addressed by Sanov in 1958 [82], using the notion of relative entropy.

Definition 2. The relative entropy of a probability measure $P \in \mathcal{P}(E)$ w.r.t R is defined by:

$$H(P|R) = \begin{cases} \mathbb{E}_P\left(\log \frac{\mathrm{d}P}{\mathrm{d}R}\right) & \text{if } P \ll R, \\ +\infty & \text{if not,} \end{cases}$$

where $\frac{\mathrm{d}P}{\mathrm{d}R}$ is the Radon-Nikodym derivative of P w.r.t R. For any measure R The function $H(\cdot|R)$ is convex and lower semi-continuous. Moreover, it vanishes only on $P = R$, and for this reason is often used as a pseudo-distance in statistics (but it is not symmetric).

Remark 3. Interestingly, the way we formulated the relative entropy is not limited to *static* measure on $\mathcal{P}(E)$, but can be also applied to *path* measure in $\mathcal{P}(C([0,T],E))$, where $\frac{\mathrm{d}P}{\mathrm{d}R}$ becomes the Radon-Nikodym derivative of P w.r.t R on $\mathcal{F}_\infty$ (or $\mathcal{F}_T$ if the process is considered on $[0,T]$).

We can now expose the Sanov theorem [82]:

Theorem 4. *Under the previous notation, by denoting $M_N \in \mathcal{P}(\mathcal{P}(E))$ the law on the empirical measure μ_N associated to the set of observations $(X^n)_{n\leq N}$, for every open set O and closed set C of $\mathcal{P}(E)$ for the narrow topology we have:*

$$\liminf_{N\to\infty} \frac{1}{N}\log(M_N(O)) \geq -\inf_{P\in O} H(P|R),$$
$$\limsup_{N\to\infty} \frac{1}{N}\log(M_N(C)) \leq -\inf_{P\in C} H(P|R).$$

We say that the sequence $(M_N)_{N\geq 0}$ satisfies a large deviations principle with rate function $H(\cdot|R)$.

Heuristically, for N large enough we then expect the probability that μ_N is in a small neighborhood V of a measure P to be:

$$\mathbb{P}(\mu_N \in V) \simeq e^{-NH(P|R)}.$$

Then, conditionally to a rare event, with high probability, everything happens as if each observations had been realized independently under the measure P instead of the reference measure R. The Sanov Theorem, which gives access to the rate function of a Large deviations problem, allows to approximate the probability of the rare event by solving a convex minimisation problem. Finally, returning to our original question about the reference measure R, by a change of point of view similar to that of a maximum likelihood problem, we can consider that the measure P characterizes the most probable law of the process *given* the reference process (which can be seen as a prior) and the observation $\mu_N \in V$.
Recalling that we are interested in this PhD to dynamical aspects of cell differentiation, we expect the biological observations to be measured not only at one, but at least a two timepoints

in order to get information about cell dynamics. Thus, following Remark 3, the reference measure has to be understood as the path measure of a stochastic process, and the empirical measure as the partial observation of the process at (at least) two timepoints. This motivates a deeper study of the entropy minimisation problem in such situation, which is called the Schrödinger problem.

The Schrödinger problem

In this section, we introduce some general results about the Schrödinger problem and its link with Large deviations of stochastic processes.

We recall that we denote R a reference path measure in $\mathcal{P}(C([0,T],E))$. R then characterizes the law of a stochastic process between 0 and $T>0$ with values in E. The joint law on the initial and final position is:

$$R_{0T} := (X_0 \times X_T)_{\#} R \in \mathcal{P}(E \times E). \tag{1.3}$$

We consider the Schrödinger problem:

$$\inf_{\pi \in \Pi(\mu,\nu)} H(\pi | R_{0T}), \tag{1.4}$$

that we denote $\mathrm{Sch}(R_{0T};\mu,\nu)$. $\Pi(\mu,\nu)$ denotes the set of probability measures $\pi \in \mathcal{P}(E\times E)$ with marginals $\mu \in \mathcal{P}(E)$ and $\nu \in \mathcal{P}(E)$ at times 0 and T, *i.e* satisfying:

$$\begin{cases} \pi(dx \times E) = \mu(dx), \\ \pi(E \times y) = \nu(dy). \end{cases}$$

Remark 5. Here, we define $\mathrm{Sch}(R_{0T};\mu,\nu)$ as the optimal value of our problem. However, with an abusive terminology, we will refer to the minimizer of the r.h.s. of (1.4) as "the solution of $\mathrm{Sch}(R_{0T};\mu,\nu)$". More generally, we will call "the problem $\mathrm{Sch}(R_{0T};\mu,\nu)$" the optimization problem consisting in computing the value $\mathrm{Sch}(R_{0T};\mu,\nu)$.

From the Sanov theorem 4, the Schrödinger problem as an interpretation in terms of large deviations. To see this, let us build an example. Consider that we are working on the tore $\mathbb{T}^d$. Consider a large number of particles that we choose independently under the law of a Brownian motion on $\mathbb{T}^d$, at times 0 and 1, with an initial random uniform position. Under these hypotheses, the law of large numbers ensures that with very high probability, the initial and final distributions of these particles are very close to be uniform. Conditionally to the rare event that they are close to two non-uniform distributions that we denote μ and ν, the Sanov theorem ensures that with very high probability, the behaviour of the system is the same as if all the particles had been chosen according to a joint probability law P^*_{01}, which is the solution of the Schrödinger problem $\mathrm{Sch}((X_0 \times X_1)_{\#}B;\mu,\nu)$ (where B denotes the path measure of the Brownian motion on $\mathbb{T}^d$).

Interestingly, it is possible from this joint law to find an optimal path measure P^* on $[0,1]$, characterizing (with very high probability) the law of the Brownian motion conditionally to the observations. This point of view was precisely Schrödinger's original ones, who aimed to understand the law of a Brownian motion at $t=\frac{1}{2}$ conditionally to two temporal marginal constraints [85]. For this, we have to introduce the so-called dynamical Schrödinger problem:

$$\inf_{P \in \mathcal{M}^+(C([0,T],E))} \left\{ H(P|R) \;\middle|\; X_{0\#}P = \mu, X_{T\#}P = \nu \right\}, \tag{1.5}$$

that we still denote $\mathrm{Sch}(R;\mu,\nu)$ when there is no ambiguity on the space on which the measure R is defined. The entropy has been characterized in that case in Remark 3.

Before considering the link between the two problems (normal and dynamical), we introduce a lemma which will be used afterwards:

Lemma 6. *Let X be a random variable with values in E. With the previous notations, we have:*

$$H(P|R) = H(X_\# P|X_\# R) + \mathbb{E}_P\left(H(P^X|R^X)\right),$$

where P^X et R^X are the conditional probabilities $P(\cdot|X)$ and $R(\cdot|X)$ respectively.

This is a simple consequence of the additive property of the relative entropy, and the proof can be found in [52] (Theorem 1.6). Typically, we can take $X = X_0$ the initial position of a process $(X_t)_{t\in[0,T]}$: $X_\# P$ and $X_\# R$ then denotes the initial law of the process under P and R respectively.

Denoting R^{xy} the bridge between x and y, *i.e* the measure defined by:

$$R^{xy}(\cdot) := R(\cdot|X_0 = x, X_1 = y),$$

we have the following theorem, the proof of which is based on Lemma 6 and can be found in [27]:

Theorem 7. *The Schrödinger problems (1.4) and (1.5) admit at most one solution, which are denoted respectively $\pi^* \in \mathcal{P}(E \times E)$ and $P^* \in \mathcal{P}(C([0,T], E))$ whenever they exist. They are characterized by the relation:*

$$P^*(\cdot) = \int\int R^{xy}(\cdot)\pi^*(dx, dy). \tag{1.6}$$

Moreover, we have $P^*_{01} = \pi^*$, and P^* shares the same bridges as R:

$$\forall x, y \in E,\ P^{*xy} = R^{xy}.$$

Finally, the costs associated to the two problems (defined as the objective values) are equal:

$$H(P^*|R) = H(\pi^*|R_{01}).$$

The equality of the bridges P^{*xy} and R^{xy} is consistent with Sanov's point of view, as we expect that the optimal path measure conditionally to the observations is the same than the reference's one.

The uniqueness of the solutions is straightforward due the strict convexity of the problems. Moreover, the link between the two problems stated by (1.6) ensures that it will be enough to consider conditions of existence for the non-dynamical Schrödinger problem (which will be the subject of Chapter 5 in a discrete setting), the solution of the dynamical one following from it.

The dynamical Schrödinger problem is then interesting as it allows to characterize the law of a stochastic process conditionally to partial observations (at given times), and a reference measure. Moreover, although it seems a priori complicated to solve in its dynamical formulation, the solution is completely determined by a coupling solving the non-dynamical Schrödinger problem, which may be simpler to solve. We are now going to detail how this theory is linked to the optimal transport (OT) theory, which has drawn a lot of attention these last 20 years, through a second type of Large deviations analysis on the reference process.

A second Large deviations principle for a stochastic process and its link with the relative entropy

In addition to the Large deviations principle in the number of observations, stated by the Sanov theorem, there exists another type of Large deviations principle for stochastic processes, which is related to the level of noise characterizing the system. It has been largely developed within the context of SDEs [29, 23]: in that case, a noise coefficient $\sqrt{\varepsilon}$ scales the diffusion coefficient, and we can then build a sequence of stochastic processes $((X_t^\varepsilon)_{t\in[0,T]})_\varepsilon$. We say that the sequence $((X_t^\varepsilon)_{t\in[0,T]})_\varepsilon$ satisfies a Large deviations principle if there exists a lower semi-continuous function $J_T : càdlàg([0,T],\mathbb{R}^n) \to [0,\infty]$, such that for all open set O and closed set C of $càdlàg([0,T],\mathbb{R}^n)$ for the narrow topology, we have:

$$\liminf_{\varepsilon\to 0} \varepsilon \log\left(\mathbb{P}((X_t^\varepsilon)_{t\in[0,T]} \in O)\right) \geq -\inf_{(\phi_t)_{t\in[0,T]}\in O} J_T((\phi_t)_{t\in[0,T]}),$$
$$\limsup_{\varepsilon\to 0} \varepsilon \log\left(\mathbb{P}((X_t^\varepsilon)_{t\in[0,T]} \in C)\right) \leq -\inf_{(\phi_t)_{t\in[0,T]}\in C} J_T((\phi_t)_{t\in[0,T]}).$$

The function J_T is called the rate function of the process in $[0,T]$, and the quantity $J_T((\phi_t)_{t\in[0,T]})$ is called the cost of the trajectory $(\phi_t)_{t\in[0,T]}$. In particular, we say that the rate function has the form of an action if the cost of any piecwise differentiable trajectory $\phi_t)_{t\in[0,T]}$ can be expressed as

$$J_T(\phi) = \int_0^T L(\phi(t), \dot{\phi}(t))dt, \tag{1.7}$$

where L is a convex lower semi-continuous function in the phase space, called the *Lagrangian* of the system.

Remark that the link between this type of Large deviations principle and Large deviation principle of random variables has been studied by Feng and Kurtz [26], and we refer to [2] for a simple exposition of these ideas. In particular, in analogy with Varadhan's lemma, which relates large deviations for sequences of random variables to the asymptotic behaviour of functionals of the form $\frac{1}{n}\log\mathbb{E}(e^{nf(X_n)})$, they extended the Fleming's approach that relates the rate function J_T to the the asymptotic behaviour of the semi-group $V^\varepsilon(T)f(x) = \varepsilon\log\mathbb{E}\left(e^{\frac{f(X_T^\varepsilon)}{\varepsilon}}|X_0^\varepsilon = x\right)$.

Interestingly, this second type of large deviations principle is still related to the entropy by the notion of Γ-convergence. We begin by recalling the definition of this type of convergence in first-countable spaces:

Definition 8. In a first-countable space E, the sequence of functional $F_\varepsilon : E \to \mathbb{R}$ is said to Γ-converge to $F : E \to \mathbb{R}$ if:

1. For every sequence X_ε in E such that $x_\varepsilon \to x$ as $\varepsilon \to 0$:

$$F(x) \leq \liminf_{\varepsilon\to 0} F_\varepsilon(x_\varepsilon),$$

2. For every x in E, there exists a sequence $x_\varepsilon \to x$ as $\varepsilon \to 0$ such that:

$$F(x) \geq \limsup_{\varepsilon\to 0} F_\varepsilon(x_\varepsilon).$$

In plain words, the first condition means that F provides an asymptotic common lower bound for the F_n, and the second condition that this lower bound is optimal. As the Skorokhod space is first-countable (the space of trajectories which are continuous to the right, with a limit to the left) the definition holds for the stochastic processes that are going to be of interest in this manuscript.

The mathematical details concerning the Γ-convergence approach of large deviations for stochastic processes are beyond the scope of this thesis, but it is interesting to have in mind that it has been recently shown [59] that there exists a large deviations principle for a sequence of stochastic processes $((X_t^\varepsilon)_{t\in[0,T]})_\varepsilon$ (with corresponding sequence of law $(R^\varepsilon)_\varepsilon$), with rate function J_T, iff:

$$\Gamma - \lim_{\varepsilon\to 0} H(P|R^\varepsilon) = \mathbb{E}_P\left(J_T((X_t)_{t\in[0,T]})\right), \tag{1.8}$$

where $H(P|R^\varepsilon)$ is then seen as a function of P.

Large deviations and optimal transport theory

We now introduce the Monge-Kantorovitch problem, which is the starting point of the OT theory, and we detail the link with what we previously exposed. Even if very good introductions to these notions can be already found elsewhere, we believe that a simple (and sometimes not completely rigorous) presentation of these results is relevant in this introduction as it highlight what should be done before considering a mechanistic approach of OT applied to biological systems.

The idea of optimal transport, initiated by Monge and Kantorovich in the beginning of the 19^{th} century, is to transport in the most economical way some mass between two prescribed distributions μ and ν. We consider $E := \mathbb{R}^n$ for simplicity, and we take as for the Schrödinger problem $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$. We also introduce a cost function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$. Then the Monge-Kantorovich problem can be written:

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} c(x,y)\pi(dx,dy). \tag{1.9}$$

In particular, when c is the quadratic cost ($c(x,y) = ||x-y||_2^2$), the quantity seen as a function of μ and ν is a distance on the space $\mathcal{P}(\mathbb{R}^n)$ called the Wasserstein distance W_2. When $c(x,y) = ||x-y||_1$, it is sometimes called the *Earth mover distance* (EMD) and is often used by bioinformaticians for comparing empirical distributions.
This problem also admits a dynamical formulation [61], which takes into account the whole trajectory of the system and not only the initial and final position and can be written:

$$\inf_{(Z_t)_{t\in[0,T]}} \Big\{ \int_{\mathbb{R}^n} C((Z_t(x))_{t\in[0,T]})\mu(dx) \,|\, T_0 = Id,\ Z_{T\#}\mu = \nu \Big\}, \tag{1.10}$$

where C is a non-negative function on $C([0,T],\mathbb{R}^n)$. These formulations are equivalent provided that the functions c and C satisfy the relation for all $x, y \in \mathbb{R}^n$:

$$c(x,y) = \inf_{(\phi_t)_{t\in[0,T]}} C((\phi_t)_{t\in[0,T]}) \,|\, \phi_0 = x, \phi_T = y,$$

when we identify the value of ϕ_t to the function $Z_t(\phi_0)$ for all $t \in [0,T]$.

For example, when c is the quadratic cost, the corresponding functional is:

$$C((\phi_t)_{t\in[0,T]}) = \int_0^T \frac{\dot{\phi}_t^2}{2} dt. \tag{1.11}$$

In this case, an important result established by Brenier in the 1980's is that whenever μ, ν are measures with second order moment finite, and μ is absolutely continuous w.r.t the Lebesgue measure, then the optimal coupling π is given by:

$$\pi = (Id \times \nabla\phi)_{\#}\mu,$$

where ϕ is a convex function such that $\nabla\phi_{\#}\mu = \nu$, called the *Brenier's map*. In the equivalent dynamical formulation, the solution is given by $(Z_t)_{t\in[0,T]}$ such that for all $0 < t < T$:

$$Z_{t\#}\mu = ((1-t)Id + t\nabla\phi)_{\#}\mu.$$

In this quadratic case, an other important result established by Benamou and Brenier in [10] states that the objective value of the OT problem (1.9) is equal to the objective value of the following optimization problem:

$$\inf_{\rho,v}\Big\{\int_{\mathbb{R}^n}\int_0^T ||v(t,x)||^2\rho(dt,dx)\,|\,\rho(0)=\mu,\ \rho(T)=\nu,\ \frac{\mathrm{d}\rho}{\mathrm{d}t}+\nabla(\rho v)=0\Big\}. \tag{1.12}$$

We observe that the quantity which is minimized in (1.12) is related to the functional C appearing in the Mc-Cann formulation: if ρ is interpreted as a time-evolving probability distribution in $\mathbb{R}^n$ associated to a stochastic process of measure P, the quantity which is minimized is precisely $\mathbb{E}_P\left(C((X_t)_{t\in[0,T]})\right)$. This form suggests that C should be related to the rate function J on the right-hand side of the formula (1.8).
It can be shown that the Benamou-Brenier formulation is precisely the Gamma-limit of the entropy minimisation problem when the reference measure is the one of a Brownian motion, for which the rate function corresponds well to the function C defined for the quadratic case by (1.11) [50].
More generally, it is common to talk about the *Benamou-Brenier* formulation of a problem when the latter is shown to be equivalent to the minimization of a quantity of the form $\mathbb{E}_P\left(C((X_t)_{t\in[0,T]})\right)$.

With these results in mind, we can now observe that the quantity on the right-hand side of (1.8) corresponds in fact to the one appearing in the Benamou-Brenier formulation of an OT problem. In particular, if the LDP has the form of an action, in analogy with the quadratic case, the cost function of the associated OT problem should be precisely defined for all $x, y \in \mathbb{R}^n$ by:

$$c(x,y) = \inf_{(\phi_t)_{t\in[0,T]}}\{\int_0^T L(\phi_t,\dot{\phi}_t)dt\,|\,\phi_0 = x, \phi_1 = y\},$$

which corresponds to the variational problem appearing for characterizing an optimal cost function in the theory of Large deviations (see Chapter 2 for more details).

All the mathematical theory that we presented in this section are related to stochastic processes. Before to see how they can be used to address the fundamental biological questions presented in the introduction, we are going to present the stochastic processes that we will consider throughout the manuscript, and the analysis of which will be at the core of the projects.

1.2 Mechanistic models of gene expression

In this chapter, we detail the different models that are used throughout the manuscript. All of them are derived, in a certain limit, from a mechanistic model describing the stochastic dynamics of promoters, mRNAs and proteins within a single cell under the action of a GRN. It has been previously developed by Ulysse Herbach in its PhD project [36].

1.2.1 Mechanistic model for one gene within a single cell

The model which is going to be used throughout this manuscript is based on a hybrid version of the well-established two-state model of gene expression [45], [74], including both mRNA and protein production [87]. A gene is described by the state of a promoter, which can be $\{on, off\}$.

If the promoter is *on*, mRNAs will be transcripted with a rate s_m and degraded with a rate d_m. If it is *off*, only mRNA degradation occurs. Translation of mRNAs into proteins happens regardless of the promoter state at a rate s_p, and protein degradation at a rate d_m. A gene is described by the state of the promoter, which can be $\{on, off\}$. k_{on} and k_{off} denote the exponential rates of transition between the states *on* and *off* (see Figure 1.1).

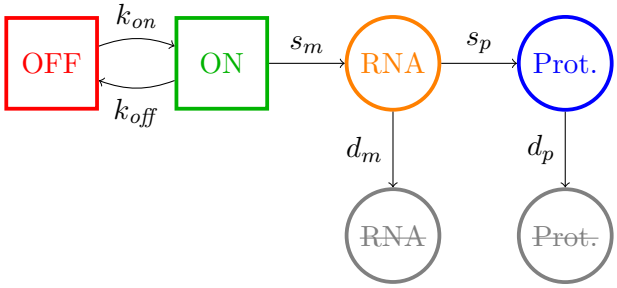


Figure 1.1: The two-states model of gene expression [38], [74].

This model can be expressed as a Markov chain process [36] on the 3 random variables describing the promoter state E, mRNA levels M and protein levels P, respectively, but in practice it is not necessary to keep a discrete description for M and P, which are abundant species and without conservation relationships. Indeed, quantitative experiments suggest that the creation and degradation parameters typically verify $s_m \gg d_m$ and $s_p \gg d_p$ [86]. In this regime, the scale of mRNA and protein levels is large enough to neglect their molecular noise, considering them as continuous quantities that follow the differential equations given by the classical mass action law. We obtain the following model, that belongs to the class of piecewise deterministic Markov processes (PDMP) [9]:

$$\begin{cases} E(t) : 0 \xrightarrow{k_{on}} 1, 1 \xrightarrow{k_{off}} 0, \\ M'(t) = s_m E(t) - d_m M(t), \\ P'(t) = s_p M(t) - d_p P(t), \end{cases} \tag{1.13}$$

$E(t), M(t), P(t)$ denote respectively the promoter state, and the mRNA and protein concentration in the cell at time t. We used here the notations of Herbach et al. [38], which were themselves inspired from the work of Rudnicki [80]: the arrows between 0 and 1 express the fact that the stochastic process $(E_t)_{t\geq 0}$ is a Markov chain with discrete space $\{0, 1\}$, continuous in time, the transitions of which follow exponential laws of rates k_{on} and k_{off}.

It is worth noticing that this model has been shown to be exactly the limit of the Markov chain described by Herbach [36], in the limit regime $s_m \gg d_m$ and $s_p \gg d_p$ [20]. In particular, this ensures that the model (1.13) is well defined as a Markov Feller process.

1.2.2 Mechanistic model for n genes in interaction within a single cell

The key idea for studying a GRN is to embed this two-states model into a network. Still denoting by n the number of genes, the vector (E, M, P) describing the process is then of dimension $3n$. The jump rates for each gene i are expressed in terms of two specific functions $k_{on,i}$ and $k_{off,i}$. To take into account the interactions between the genes, we consider that for all $i = 1, \cdots, n$, $k_{on,i}$ is a function which depends on the full vector P via the GRN, represented by a matrix θ of size n. We denote these functions $k^{\theta}_{on,i}$ and assume that $k^{\theta}_{on,i}$ is upper and lower bounded by a positive constant for all i. The function is chosen such that if gene i activates gene j, then $\partial_{P_i} k_{on,j} \geq 0$. For the sake of simplicity, and because it is experimentally observed that the time spent by a promoter on state 1 is poorly regulated w.r.t the frequency of the transition between the states 0 and 1, we consider that $k_{off,i}$ does not depend on the protein levels. We obtain a

system of n PDMP coupled by the jump rate functions $k^{\theta}_{on,i}$

$$\forall i = 1, \cdots, n : \begin{cases} E_i(t) : 0 \xrightarrow{k^{\theta}_{on,i}(P(t))} 1, 1 \xrightarrow{k_{off,i}} 0, \\ M_i'(t) = s_{m,i} E_i(t) - d_{m,i} M(t), \\ P_i'(t) = s_{p,i} M_i(t) - d_{p,i} P_i(t). \end{cases} \tag{1.14}$$

As a system of n Markov Feller processes, this is also a Markov Feller process.

The precise form of the functions $k_{on,i}$ will be detailed afterwards. Note that a model of chromatin dynamics around the promoters has been built in Herbach et al. [38] and used for deriving a mechanistic form of these functions. It will be used in Chapter 2, and then simplified into a multivariate sigmoidal function in the next chapters.

1.2.3 Simplified model describing proteins dynamics

A scaling analysis allows to simplify this model. Indeed, degradation rates play a crucial role in the dynamics of the system. The ratio $\frac{d_{m,i}}{d_{p,i}}$ controls the buffering of promoter noise by mRNAs and, since it is observed in practice that $k_{off,i} \gg k_{on,i}$ [69, 79], the ratio $\frac{k_{on,i}}{d_{m,i}}$ controls the buffering of mRNA noise by proteins. In line with several experiments [4, 53], we consider that mRNA levels evolve rapidly in regards to protein levels dynamics, *i.e* $\frac{d_{m,i}}{d_{p,i}} \gg 1$ with $\frac{k_{on,i}}{d_{m,i}}$ fixed. The correlation between mRNAs and proteins produced by the gene is then very small, and the model can be reduced by removing mRNA and making proteins directly depend on the promoters. A mathematical analysis of this statement can be found in [36], and we refer to [105] for a rigorous proof in the case of a model of gene expression close to this one.

We then obtain the following simplification: if the state of the promoter is *on*, mRNAs are transcripted and translated into proteins, which are considered to be produced at a rate s. If the state of the promoter is *off*, only degradation of proteins occurs at a rate d (see Figure 1.2).

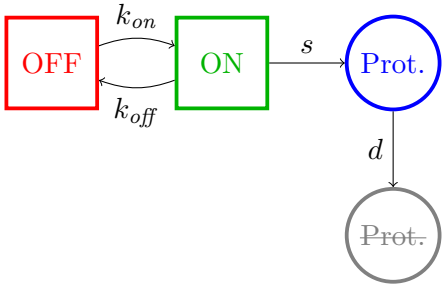


Figure 1.2: Simplified two-states model of gene expression.

Finally, the parameters s_i can be removed by a simple rescaling of the protein concentration P_i for every gene i by its equilibrium value when $E_i = 1$ (see [38] for more details). We obtain a reduced dimensionless PDMP system modeling the expression of n genes in a single cell:

$$\forall i = 1, \cdots, n : \begin{cases} E_i(t) : 0 \xrightarrow{k^{\theta}_{on,i}(X(t))} 1, 1 \xrightarrow{k_{off,i}} 0, \\ X_i'(t) = d_i(E_i(t) - X_i(t)). \end{cases} \tag{1.15}$$

Here, $X(t)$ describes the protein vector in the renormalized gene expression space $\Omega := (0,1)^n$ and $E(t)$ the promoters state in $P_E := \{0,1\}^n$, at time t. This model will be used in Chapter 2.

1.2.4 Mechanistic model in the bursty regime and simplification

We now consider the so-called bursty regime of this model, when $k_{on} \ll k_{off}$, which corresponds to the experimentally observed situation where active periods are short but characterised by a high transcription rate, thereby generating bursts of mRNA [69], [79], [94]. We describe the random times at which these bursts occur by an exponential law of parameter k_{on}, and their random intensity by an exponential law of parameter k_{off}/s_m (see Figure 1.3).

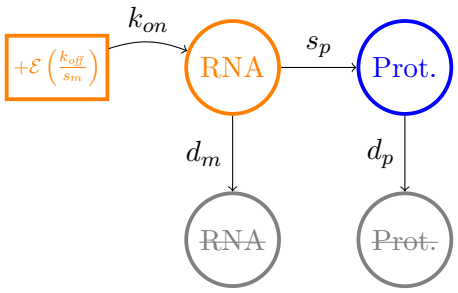


Figure 1.3: Approximation of the two-states model of gene expression in the bursty regime.

As for the model with promoters, neglecting the molecular noise associated to mRNA and protein levels and adding interactions between genes through the functions $k^{\theta}_{on,i}$, we obtain the following mathematical description of the model:

$$\forall i = 1, \cdots, n : \begin{cases} M_i(t) \xrightarrow{k^{\theta}_{on,i}(P(t))} M_i(t) + \mathcal{E}\left(\frac{k_{off,i}}{s_{m,i}}\right), \\ M_i'(t) = -d_{m,i} M_i(t), \\ P_i'(t) = s_{p,i} M_i(t) - d_{p,i} P_i(t). \end{cases} \tag{1.16}$$

Using the same arguments as for the model (1.15), we can simplify this model by removing mRNAs. In that case, we remark that it is no more necessary to rescale protein concentrations, as the creation rates $s_{p,i}$ only appear in the exponential law characterizing the jumps. We obtain:

$$\forall i = 1, \cdots, n : \begin{cases} P_i(t) \xrightarrow{k^{\theta}_{on,i}(P(t))} P_i(t) + \mathcal{E}\left(c_i\right), \\ P_i'(t) = -d_i P_i(t), \end{cases} \tag{1.17}$$

where we define $c_i = \frac{k_{off,i} d_{m,i}}{s_{m,i} s_{p,i}}$. This model will be used in Chapters 3-6.

1.2.5 Simulations and stationary distributions of the models in case of constant rate functions

When the rate functions $k^{\theta}_{on,i}$ for every gene i are constant (then the parameter θ has no effect and can be removed from the notations), the models correspond to a set of n independent PDMP. In that case, some of the marginal stationary distributions are known. For the models taking into account mRNAs and proteins, the marginals on mRNAs of the stationary distribution are known, and correspond to Beta and Gamma distributions for the models (1.14) and (1.16) respectively. For the simplified models without mRNAs, the marginal on proteins of the stationary distributions are known, corresponding also to Beta and Gamma distributions for the models (1.15) and (1.17), respectively. These models are thus compatible with real single-cell data: indeed Beta and Gamma distributions, or multimodal mixtures of them, are known to describe accurately single-cell data [4, 16, 58].
The form of these stationary distributions and the reference where the proofs can be found are summarized in Table 1.1. We also illustrate a stochastic trajectory associated to each of these models in Figure 1.4.

Model	Stationary distribution	References
(1.14)	$M \sim \frac{s_{m,i}}{d_{m,i}}\text{Beta}\left(\frac{k_{on,i}}{d_{m,i}}, \frac{k_{off,i}}{d_{m,i}}\right)$	[21, 37]
(1.15)	$P \sim \frac{s_{m,i}s_{p,i}}{d_{m,i}d_{p,i}}\text{Beta}\left(\frac{k_{on,i}}{d_{p,i}}, \frac{k_{off,i}}{d_{p,i}}\right)$	[21, 38]
(1.16)	$M \sim \text{Gamma}\left(\frac{k_{on,i}}{d_{m,i}}, \frac{k_{off,i}}{d_{m,i}}\right)$	[57, 38]
(1.17)	$P \sim \text{Gamma}\left(\frac{k_{on,i}}{d_{p,i}}, \frac{d_{m,i}k_{off,i}}{s_{m,i}s_{p,i}}\right)$	[30, 56]

Table 1.1: Known stationary distributions for models described in this section, with M and P denoting mRNA and protein levels. The stationary distribution of P in complete models (1.14) and (1.16) is still lacking in terms of analytical results, motivating the construction of the reduced models. This table is inspired from Herbach [35].

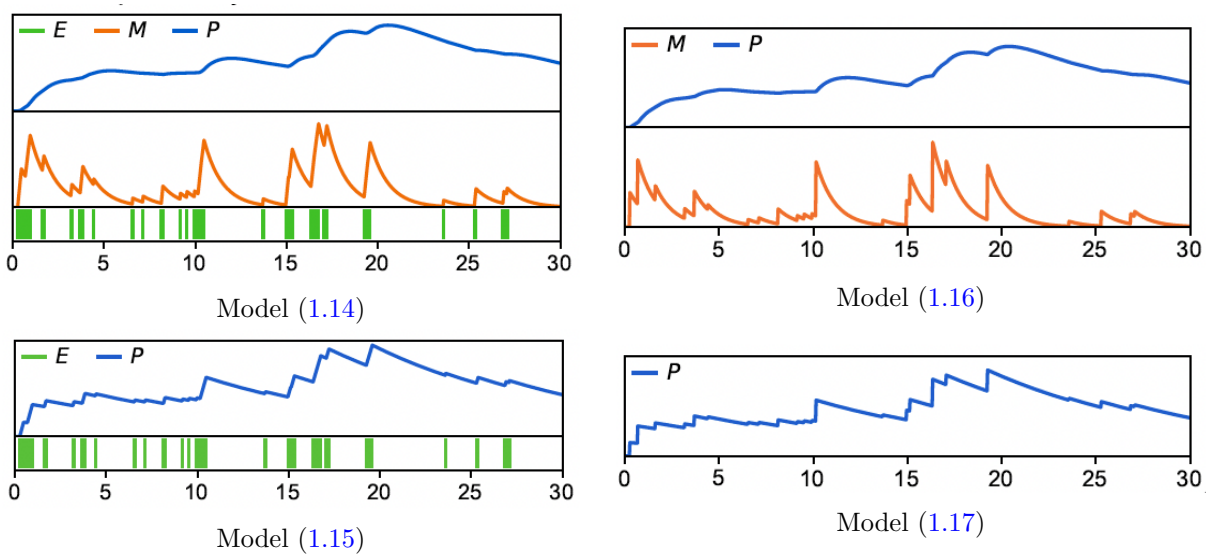


Figure 1.4: Sample time trajectories for the stochastic gene expression models described previously. Promoter activity, mRNA levels and protein levels are denoted by E, M and P, respectively. Figure from Herbach [35].

1.2.6 Master equation for the models describing proteins dynamics

As all the processes characterized by the models previously described are Markov Feller processes, they are characterized by their infinitesimal generator, and there exists a probability distribution characterizing the evolution of the process which is solution to the master equation associated to the generator. We describe in this section the master equations associated to the models (1.15) and (1.17) that are going to be used in the different parts of the manuscript. The generator of the model (1.15) appears as the limit of the generator of the pure jump process characterizing the dynamics, described in [36], which has been proved in Crudu et al. [20]. The generator of the model (1.17) is the limit of the generator of the model (1.15) when $k_{off,i} \gg k^{\theta}_{on,i}$ with $k_{off,i}/s_{p,i}$ fixed. We provide a complete proof of this convergence result in Chapter 6, strongly inspired by [20].

Master equation of the PDMP model with promoters

As $\text{card}(P_E) = 2^n$, we can write the joint probability density $\rho(t, e, x)$ of (E_t, X_t) as a 2^n-dimensional vector $\rho(t, x) = (\rho(t, e, x))_{e \in P_E} \in \mathbb{R}^{2^n}$. The master equation on u can be written:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t}(t, x) + \sum_{i=1}^{n} \partial_{x_i} \left(F_i(x)\rho(t, x)\right) = \sum_{i=1}^{n} K_i(x)\rho(t, x). \tag{1.18}$$

For all $i = 1, \cdots, n$, for all $x \in \Omega$, $F_i(x)$ and $K_i(x)$ are matrices of size 2^n. Each F_i is diagonal, and the term on a line associated to a promoter state e corresponds to the drift of gene i: $d_i(e_i - x_i)$. K_i is not diagonal: each state e is coupled with every state e' such that only the coordinate e_i changes in e, from 1 to 0 or conversely. Each of these matrices can be expressed as a tensorial product of $(n-1)$ two-dimensional identity matrices with a two-dimensional matrix corresponding to the operator associated to an isolated gene:

- $F_i(x) = I_2 \otimes \cdots \otimes \underbrace{F^{(i)}(x)}_{i^{th} \text{ position}} \otimes \cdots \otimes I_2$
- $K_i(x) = I_2 \otimes \cdots \otimes \underbrace{K^{(i)}(x)}_{i^{th} \text{ position}} \otimes \cdots \otimes I_2$,
- $F^{(i)}(x) = \begin{pmatrix} -d_i x_i & 0 \\ 0 & d_i(1-x_i) \end{pmatrix}$
- $K^{(i)}(x) = \begin{pmatrix} -k^\theta_{on,i}(x) & k_{off,i}(x) \\ k^\theta_{on,i}(x) & -k_{off,i}(x) \end{pmatrix}$.

For the sake of clarity, we detail this tensorial expression (1.18) for a two-dimensional network. The general form for the infinitesimal operator can be written:

$$L\rho(t,e,x) = \langle F(e,x), \nabla\rho(t,e,x)\rangle + \sum_{e' \in P_E} Q(e,e')(x)\rho(t,e',x)$$

where F is the vectorial flow associated to the PDMP and Q the matrix associated to the jump operator. A jump between two promoters states e, e' is possible only if there is exactly one gene for which the promoter has a different state in e than in e': in this case, we denote $e \sim e'$.
We have, for any x: $F(e,x) = (d_0(e_0 - x_0), \cdots, d_n(e_n - x_n))^T$. Then, for all $e \in P_E$, the infinitesimal operator can be written:

$$L\rho(t,e,x) = \sum_{i=1}^{n} F_i(e,x)\partial_{x_i}\rho(t,e,x) + \sum_{\{e'|e' \sim e\}} \left(k^\theta_{on,i}(x)\delta_{e_i=0} + k_{off,i}\delta_{e_i=1}\right)\left(\rho(t,e',x) - \rho(t,e,x)\right).$$

For a two-dimensional process ($n = 2$), there are four possible configurations for the promoter state: $e_{00} = (0,0)$, $e_{01} = (0,1)$, $e_{10} = (1,0)$, $e_{11} = (1,1)$. It is impossible to jump between the states e_{00} and e_{11}. If we denote $\rho(t,x)$ the four-dimensional vector: $(\rho(t,e,x))_{e \in P_E}$, we can write the infinitesimal operator in a matrix form:

$$\begin{aligned} L\rho(t,x) = &\underbrace{\begin{pmatrix} -d_1x_1 & 0 & 0 & 0 \\ 0 & -d_1x_1 & 0 & 0 \\ 0 & 0 & d_1(1-x_1) & 0 \\ 0 & 0 & 0 & d_1(1-x_1) \end{pmatrix}}_{F_1(x)} \begin{pmatrix} \partial_{x_1}\rho_{e_{00}}(t,x) \\ \partial_{x_1}\rho_{e_{01}}(t,x) \\ \partial_{x_1}\rho_{e_{10}}(t,x) \\ \partial_{x_1}\rho_{e_{11}}(t,x) \end{pmatrix} + \\ &\underbrace{\begin{pmatrix} -d_2x_2 & 0 & 0 & 0 \\ 0 & d_2(1-x_2) & 0 & 0 \\ 0 & 0 & -d_2x_2 & 0 \\ 0 & 0 & 0 & d_2(1-x_2) \end{pmatrix}}_{F_2(x)} \begin{pmatrix} \partial_{x_2}\rho_{e_{00}}(t,x) \\ \partial_{x_2}\rho_{e_{01}}(t,x) \\ \partial_{x_2}\rho_{e_{10}}(t,x) \\ \partial_{x_2}\rho_{e_{11}}(t,x) \end{pmatrix} + \\ &\underbrace{\begin{pmatrix} -k_{on,1}(x) & 0 & k_{on,1}(x) & 0 \\ 0 & -k_{on,1}(x) & 0 & k_{on,1}(x) \\ k^\theta_{off,1} & 0 & -k^\theta_{off,1} & 0 \\ 0 & k^\theta_{off,1} & 0 & -k^\theta_{off,1} \end{pmatrix}}_{Q_1(x)} \begin{pmatrix} \rho_{e_{00}}(t,x) \\ \rho_{e_{01}}(t,x) \\ \rho_{e_{10}}(t,x) \\ \rho_{e_{11}}(t,x) \end{pmatrix} + \\ &\underbrace{\begin{pmatrix} -k_{on,2}(x) & k_{on,2}(x) & 0 & 0 \\ k^\theta_{off,2} & -k^\theta_{off,2} & 0 & 0 \\ 0 & 0 & -k_{on,2}(x) & k_{on,2}(x) \\ 0 & 0 & k^\theta_{off,2} & -k^\theta_{off,2} \end{pmatrix}}_{Q_2(x)} \begin{pmatrix} \rho_{e_{00}}(t,x) \\ \rho_{e_{01}}(t,x) \\ \rho_{e_{10}}(t,x) \\ \rho_{e_{11}}(t,x) \end{pmatrix}. \end{aligned}$$

We remark that each of these matrices can be written as a tensorial product of the corresponding two-dimensional operator with the identity matrix:

- $F_1(x) = F^{(1)}(x) \otimes I_2$
- $F_2(x) = I_2 \otimes F^{(2)}(x)$
- $F^{(i)}(x) = \begin{pmatrix} -d_i x_i & 0 \\ 0 & d_i(1-x_i) \end{pmatrix}$
- $Q_1(x) = Q^{(1)}(x) \otimes I_2$
- $Q_2(x) = I_2 \otimes Q^{(2)}(x)$
- $Q^{(i)}(x) = \begin{pmatrix} -k^\theta_{on,i}(x) & k^\theta_{on,i}(x) \\ k_{off,i} & -k_{off,i} \end{pmatrix}$.

The master equation (1.18) is obtained by taking the adjoint operator of L:

$$\frac{\mathrm{d}u}{\mathrm{d}t}(t,x) = L^*\rho(t,x) = -\sum_{i=1}^{2} \partial_{x_i}(F_i u)(t,x) + \sum_{i=1}^{2} K_i \rho(t,x)$$

where $K(x) = Q^T(x)$ is the transpose matrix of Q.

Master equation of the bursty model

The master equation on the probability density $\rho(t,\cdot)$ of the bursty model (1.17), describing only proteins, associated to a GRN θ appears as an integro-differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t,x) = \sum_{i=1}^{n} \left[\partial_{x_i}\left[d_i x_i \rho(t,x)\right] + \int_0^{x_i} k^\theta_{on,i}(x - hb_i)\rho(t, x-hb_i) c_i e^{-c_i h} dh - k^\theta_{on,i}(x)\rho(t,x) \right], \tag{1.19}$$

where for all i, b_i is a vector of size n with only zero entries except on the i^{th} position.

1.3 Existing probabilistic methods for analyzing gene expression data

In this section, we clarify how the biological questions presented in the introduction will be articulated with the mathematical questions. We start by presenting the type of data with which we are going to work and from which we will thus try to address these questions. We then present, in a non-exhaustive way, some methods developed these last few years for exploiting such data. We separate the presentation of methods aiming to infer a **GRN** from the ones aiming to infer a **landscape**. It is important to keep in mind that, as the landscape is considered to be mainly shaped by a GRN, these methods follow implicitly a similar goal provided that there exists a notion of landscape associated to a GRN model. However, there are only a few methods that take a dynamical probabilistic point of view in the field of GRN inference, and a few models used for trajectories reconstruction that take into account a GRN. This explains why these two fields are generally considered as distinct. An important goal of our PhD will be to articulate them, using the mechanistic models presented in Section 1.2.

1.3.1 Gene expression data

The data that we will consider consists in the total number of mRNAs expressed by cells, also called transcriptomic data or gene expression measurements. They can be divided into population-based (when we have access to the mean expression level on a population of cells) and single-cell-based (when we have access to the expression of each cell, individually). As mentioned before, we are interested in the second type because it contains the richest information: from a probabilistic point of view, they give access to a joint probability distribution of gene expression, while the first type corresponds only to the average. Thus, mathematically, it can be considered

that such data corresponds to partial observations of the stochastic process modeling each cell of the population.
However, measurement techniques used for obtaining these data involve the physical destruction of the cells that are measured: thus, when measurements are made at several time points, for example to study how a population of cells would evolve in response to a perturbation, we do not obtain cell trajectories: the data then consists in series of time-ordered datasets of independent cells, that are called time-stamped datasets [77, 31] in the rest of the manuscript. In practice, the data that we use in this PhD correspond to count tables (or series of them in case of time-stamped datasets), of size $C \times n$, where C denotes the number of cells and n the number of genes measured. Each coordinate of a table corresponds to an integer value, corresponding to the number of mRNA fragments associated to a gene that have been measured experimentally. Remark that these count tables are themselves the result of a pre-treatment on the raw data obtained using standard bioinformatic tools, but the latter is beyond the scope of our work.

It is worth noticing that many measurement technologies allow to get different observations at the single-cell level. In addition to the class of transcriptomic data that we just described, we can mention:

- The class of epigenomic data:
  - ChIP-seq [42], that detects proteins interactions with DNA, giving access to sequences on which proteins have been fixed. We are going to use these data in Chapter 4 for assessing whether the genes interactions predicted by a GRN are supported by physical interactions or not;
  - ATAC-seq [88], that assesses genome-wide chromatin accessibility. They can also be used for detecting the binding of proteins on some sequences of the chromatin. Their interest is to have very large information about the chromatin accessibility of the genome, which can be explained by the transcriptional activity, but also to physical constraints;
  - Hi-C [95], that quantifies interactions between fragments of DNA along the genome. They are nevertheless more difficult to relate directly to the transcriptional activity than the two previous ones, but are central for understanding how information, for example about chromatin accessibility, can propagate along the genome.
- The class of proteomic data [99], which estimates the total number of proteins expressed by cells. It can be realized using flow cytometry with fluorescent markers, or mass spectrometry. In the first case, we can have access to single-cell level, but only for membrane proteins and when specific antibodies known for binding to the proteins are known. Mass spectrometry is supposed to overcome these limits, but the development of these techniques at the single-cell level are limited at the moment [44].

Finally, there now exists techniques allowing to trace the lineage of a population of cells, called CRISPR-based lineage tracing [63], allowing the identification of all progeny of a single cell, or techniques of spatial transcriptomics that give access to the transcriptomic profiles as well as the spatial position of the cells, and thus allows to question possible communication between cells. Combining these type of data for a set of cells would be of great interest for understanding the mechanisms which drive their development. This is being made possible by the advent of multiomics technologies [49, 90], that will probably revolutionise cell biology in the coming years. However, these multiomics data are very difficult to obtain, especially at multiple timepoints as we want to use (to study dynamical processes) in this PhD.

As mentioned in the introduction, transcriptomic data are difficult to analyze. Indeed, mRNA levels are highly-variable (which is mainly due to their low half-life value), and the transcriptional

bursting phenomenon makes their distribution far from Gaussian, which prevents from using many standard statistical tools used in data analyzes. They are nevertheless the most accessible at the moment, at the single-cell level, which explains why we are going to focus on these data in the following.

1.3.2 Dynamical methods for reconstructing a GRN

One of the main challenges for biologists for the ten last years is then to find relevant tools to analyze these widely available datasets. In practice, a wide range of multivariate methods have been applied to expression data and more specifically to gene network inference [41, 103, 68]. For example, one popular inference algorithm is based on a regression approach using random forests [40], and Granger causality is often used for reconstructing the interactions [72, 76]. These statistical approaches suffer from a principal drawback, underlined for example by Chan et al. [15], which is that it is very easy to infer graphs (and then GRN) from the data, it is very difficult to know if these graphs correspond to a biological reality. Indeed, there are almost no known reference networks in systems biology, and the few that are considered as such are often derived from old analyses that the new type of available data, and the associated new probabilistic paradigm, are now questioning. We should then be interested in methods that are able not only to reconstruct a GRN from experimental data, but also to simulate new data from this GRN in order to test the inferred GRN. This is the novelty of the approach developed by Bonnaffoux et al. [12], that we will developed in a more scalable way.

Let us recall that one of the main questions that we addressed in the introduction was to understand differentiation of a cell as the action of an underlying GRN driving the molecular reactions within the cell, in such a way it would be possible to reconstruct this GRN from experimental data. In our probabilistic framework, we can see the evolution of a cell as the realization of a stochastic process, and the data as partial observations of this process, $i.e$ of a sequence $(\rho_1, \cdots, \rho_n)$ when the number of cells measured at each timepoint is large enough. These goals can be now detailed in a precise mathematical framework:

- Build a model of the stochastic process characterizing differentiation configured by a GRN;
- Make appear the link between the temporal distributions and the underlying GRN;
- Find the most-likely GRN associated to a sequence of the form $(\rho_1, \cdots, \rho_n)$.

This is in fact a generalization of methods suited for population data which aimed to reconstruct the drift of the underlying ODE modeling the average trajectory of cells from its value at different timepoints. Some methods recently developed for single-cell data still used a similar framework, using the variability for improving the drift estimation [5, 60] or transforming the ODE into an SDE of the form (1.1) with constant diffusion coefficient [1].

However, while population data were not able to question the normal distribution paradigm, meaning that the variability between cells would be the effect of a high number of independent small perturbations and not central in the study of differentiation, single-cell data revealed that this variability is better described by Gamma distributions or multimodal mixtures of them [16, 58], which is known to be related to the bursty transcription of mRNAs [4, 38]. In that context, models based on ODEs or SDEs does not hold anymore, unless adapted artificial noise is added as an additional layer [24], leading to untractable analyzes. This motivated the introduction of the models presented in Section 1.2, which we have shown to be compatible with these Gamma distribution, at least when the rate functions are constant. Note that it has been also shown in Herbach [36] that the model (1.15) was able to generate mixture of Beta distributions when the rate functions were depending only on the gene that they regulated. However, neither

the temporal nor stationary distribution of these models are explicitly known when the rate functions $k^{\theta}_{on,i}$ depend on the whole protein field through a general GRN represented by a matrix θ.

The 3-step method described previously consists precisely in the approach of the algorithm HARISSA [38], when only one distribution ρ is observed and is supposed to be stationary, using the model (1.14). It has been extended more recently [35] using the model (1.16). Another numerical method called WASABI [12], which uses a divide-and-conquer approach where the problem of GRN inference is solved one gene at a time, has been developed at the same time. However, these approach suffered from some drawbacks. For HARISSA, it was accurate for small networks but computationally intractable for more than a few genes in its first version, and not precise enough in the second version, regarding other existing inference algorithms. The main uncertainty stemmed from the fact that this method was based on a so-called Hartree approximation, inspired by statistical physics, providing a parametric distribution (w.r.t a GRN) but for which there was no bound for the error characterizing the deviation to the true stationary distribution, and which was difficult to extend in the dynamical case. For WASABI, although it was able to propose relevant GRNs, it required days of computation for a GRNs with 50 genes and proposed a (potentially long) list of candidates GRNs.

1.3.3 Methods for reconstructing a landscape

We now describe methods aiming to analyze or reconstruct a landscape, independently of the notion of an underlying GRN. We emphasize that these methods could be generally used, by adding a parametric model on the functions characterizing the landscape, for GRN inference purposes as it has been presented in the previous section.

Methods for landscape analyses

Before talking about landscape reconstruction, we have to understand precisely how we can characterize a landscape associated to a model of gene expression. We recall that, as mentioned in Section 1.1.1, for a deterministic system of the form $\dot{X}_t = -\nabla V(X_t)$, the potential V provides a first natural candidate for characterizing the landscape.
In the case of a SDE with non-gradient drift like (1.1) with constant diffusion coefficient σ, the notion of potential is still natural. Indeed in that case the master equation is of the form:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}x}(F\rho) + \frac{\sigma}{2}\frac{\mathrm{d}^2\rho}{\mathrm{d}x^2}. \tag{1.20}$$

If the stationary distribution $\hat{u}$ exists, we have :

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(-F\hat{u} + \frac{\sigma}{2}\frac{\mathrm{d}\hat{u}}{\mathrm{d}x}\big) = 0 \to \big(-F\hat{u} + \frac{\sigma}{2}\frac{\mathrm{d}\hat{u}}{\mathrm{d}x}\big) = \hat{J}$$

where $\hat{J}$ is called the stationary flux. Thus, if $\hat{u}$ is non-zero everywhere, we have:

$$F = -\frac{\hat{J}}{\hat{u}} - \frac{\sigma}{2}\frac{\mathrm{d}}{\mathrm{d}x}(-\ln\hat{u}) = F_r - \nabla V,$$

where $F_r = -\frac{\hat{J}}{\hat{u}}$ can be seen as the rotational part of the drift, and $V = \frac{\sigma}{2}(-\ln\hat{u})$ the part which derives from a potential, which is then called the *quasipotential*. We can remark that he nullity of the flux depends directly on the possibility to find a stationary distribution of the form $\hat{u}(x) \sim e^{-V(x)}$. More details about the interpretation of the quasi-potential defined in this way can be found in [108].

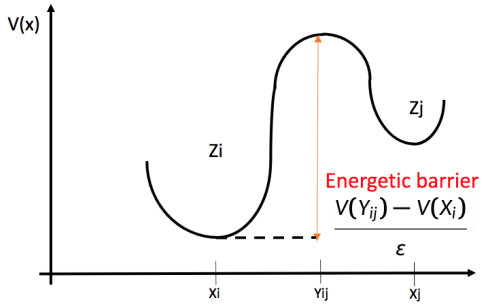


Figure 1.5: Potential landscape characterizing an energetic barrier between two basins Z_i and Z_j, corresponding to areas of high probability in the gene expression space. X_i and X_j are the attractors associated to these basins, and Y_{ij} is a saddle point at the boundary of these basins.

A general definition of a quasipotential for SDEs has been introduced by Freidlin and Wentzell [29], when the diffusion is scaled by a noise coefficient ε:

$$\mathrm{d}X_t = F(X_t)\mathrm{d}t + \sqrt{\varepsilon}\sigma(X_t)\mathrm{d}B_t, \tag{1.21}$$

Indeed for such SDEs, they showed that under some regularity conditions on the drift and the diffusion coefficient, these stochastic processes were satisfying a Large deviations principle which had the form of an action. In that situation, the Fenchel-Legendre transform of the Lagrangian,

$$H(x,p) = \sup_v \langle x, v\rangle - L(x,v),$$

is called the Hamiltonian of the system and allows to define the quasipotential of the process as the unique solution (whenever there exists) of the so-called stationary Hamilton-Jacobi equation:

$$\forall x \in \mathbb{R}^n : \; H(x, \nabla V(x)) = 0.$$

It can be proved (and it will be in Chapter 2) that when V is C^1 on $\mathbb{R}^n$, the difference of the function V characterizes the minimum of the rate function on some specific sets of trajectories, and then characterizes probabilities of realizing stochastic transitions between the basins of attraction of the deterministic system $\dot{X}_t = F(X_t)$. Note that since the work of Huang [39], these basins are generally associated to the cell types, which is a very convenient mathematical definitions of this notion (provided that cell dynamics is well described by such stochastic process). Indeed, it is consistent with the probabilistic point of view on differentiation while justifying the reproducibility and stability of the process at the functional level, as soon as the noise coefficient is small enough for the transitions between these cell types to be rare. Note that we are going to adapt this characterization for the mechanistic model (1.15) in Chapter 2.

It has also been proved that the function V corresponds to the first-order approximation in ε of the transform $-\log \hat{u}$, where $\hat{u}$ is the stationary distribution of the SDE [29]. Thus, this function V makes it possible to match the two expected definitions of potential, as it characterizes both the areas of high and low probability in the gene expression space and the optimal trajectories between the basins of attraction of the deterministic system, in the weak noise limit $\varepsilon \to 0$. When there is no regular solution V, the potential may also be defined by the minimum of the rate function on specific sets of trajectories in the gene expression space [109], in order to characterize at least the transitions between the different "potential wells", *i.e* the areas of highest probabilities, as illustrated in Figure 1.5.

This methodology has been used for characterizing the landscape associated to mechanistic stochastic models of gene expression [55, 54], but in simple cases where only one or two chemical

species were modeled. In the second reference, the model was a stochastic hybrid process close to the model (1.15), but the results were limited to the case of a very specific network. In Chapter 2, we will use Large deviations theory for reducing the PDMP model (1.15) into a coarse-grained discrete model on the cellular types.

Landscape reconstruction using probabilistic tools

In the last few years, mathematicians have rushed to develop methods for reconstructing a landscape, seen as a potential energy surface, from gene expression datasets. These tools were generally developed for reconstructing cellular dynamics in the gene expression space, under the hypothesis that cellular development are governed by such potential, in order to overcome the fundamental limitation that we do not have access to cell trajectories. They may use static data [104, 73, 106] or time-stamped datasets [84, 31].

A first interesting example is the work developed by Pearce et al. [73], who propose the following method for reconstructing state transition of a population of cells from a static dataset, seen as a stationary distribution:

1. Infer from the observations a mixture of Gaussian distributions $\hat{u}$, supposed to approximate the distribution of the dataset. The function $V = -\ln \hat{u}$ then approximates the potential characterizing the landscape associated to the observations;
2. Find the minima and maxima of this potential function and identify the transitions between two areas between two maxima to a cellular type;
3. Use the difference of potential between these minima and maxima for characterizing (up to an adequate correction) the energetic barrier between each pair of cellular type, and deduce the associated transition rates.

They obtained a Markovian transition network directly from time-independent data sampled from stationary equilibrium distributions, without introducing any underlying model. To our point of view, this approach is of interest as it uses a proxy for the stationary distribution (a Gaussian mixture), which is accessible in high dimension, to reconstruct an approximate landscape. As it will be discussed further, Chapters 2 and 3 will adopt a similar approach, but achieved in a mechanistic way. It will also overcome some limitations of this work, as the important fact that the maxima of the potential function V corresponds to areas of the gene expression space with low probability, and are then very difficult to estimate reliably.

A second example has been developed by Gao et al. in [31], where the authors reconstructed a measure of what they called the "transcriptional uncertainty" from time-course series of datasets. It corresponds to the potential $V = -\ln \hat{u}$, where $\hat{u}$ is the stationary distribution of a model of the form (1.14), when the exact stationary distribution is the one of an uncoupled system of PDMPs (*i.e* with only self-regulated genes). They applied this method for revealing that a peak of this transcriptional uncertainty is experimentally observed on several datasets during differentiation, before decreasing. Interestingly, they used this same approximation for GRN inference purposes [72], by computing temporal changes in gene expression through the distance between two consecutive timepoints of the marginal distributions before using Granger causality for reconstructing GRN causal links. Note that our developments of Chapters 2-4 will aim to overcome this approximation by reconstructing an approximation of the potential V which takes into account the underlying GRN, allowing for a better reconstruction of the landscape and a direct inference of the GRN from the approximate potential V which does not requires additional statistical tools.

Finally, we present two interesting landscape reconstruction methods have been developed under the hypothesis that cell dynamics is well modeled by a system of SDEs. Despite this restrictive

assumption, their great strength consists in taking into account the proliferation of cells in addition to the variability due to the transcriptional activity, paving the way for a broader approach of cellular variability. Note that the term proliferation has to be understood in this section as a possible creation (birth of new cells, by division) and loss (death of cells, for example by apoptosis) of mass for the observed population of cells.

The first one, which is called PBA, has been proposed in Weinreb et al. [104], and aims to reconstruct a Markov chain between a set of observed cells under the hypothesis that they are sample under the stationary distribution of a SDE with proliferation, the master equation of which is of the form (1.20) plus an addition term $C(x)u(t,x)$ where C is the proliferation rate and u the distribution of cells. They construct a graph associated to the observation using the k-nearest-neighbours method, each node being characterized by a cell and the k-neighbours of a cell being the k most similar cells. Using mathematical results on graph theory ensuring that an asymptotically exact solution of the master equation can be calculated on such nearest-neighbor graph, they estimate the the potential V of the system and define the law P of the Markov chain between the observed cells using the so-called Arrhenius formula:

$$P_{ij} \sim e^{V(X_j)-V(X_i)},$$

where X_i and X_j are the two vectors characterizing the cells i and j in the gene expression space and P_{ij} is the transition rate associated to the Markov chain on the discrete space of cells. This Markov chain allows the construction of pseudo-trajectories, as well as probabilities associated to distinct cell fates. This method is particularly interesting because the Markov chain has been shown to converge asymptotically to the underlying continuous stochastic process in the limit of an infinite number of cells. We nevertheless believe that for available datasets, we are too far from this asymptotic hypothesis to be able to predict reliably the potential function in most areas of the gene expression space: that is why we will not use directly graph theory on the observations, but rather we will reduce our mechanistic process into a Markov chain on the discrete space of cell types, identified from the observations (see Chapters 2-4).

The second one, which is called Waddington-OT, is going to be particularly of interest for us, as it will be developed in a mechanistic version (but not exactly for the same purpose) in Chapter 6. It has been developed in Schiebinger et al. [84], and mathematically refined in Lavenant et al. [47]. The aim of the method is to reconstruct most-probable trajectories of cells from time-stamped data. More precisely, the authors use the Schrödinger problem for reconstructing a serie of most-probable couplings between time-stamped datasets. They deduced a most-probable trajectory in the space of distribution with values on the gene expression space, given observations and for a given level of noise, and they predicted associated most probable trajectories of cells in the limit of small noise, using analogies with OT theory. The principle still relies on the hypothesis that cell dynamics is driven by a SDE of the form (1.21), with a gradient drift. They estimate this gradient by solving a Schrödinger problem of the form (1.4), where the reference measure is the joint law on the initial and final position associated to a simple Brownian motion. For justifying this reference measure, they prove the following theorem:

Theorem 9 (Lavenant, 2021)**.** *If $P \in \mathcal{P}(C([0,T],\mathbb{R}^n))$ characterizes a solution of the SDE 1.21 with gradient drift (with $H(P_0|Leb) < \infty$), and consider $R \in \mathcal{P}(C([0,T],\mathbb{R}^n))$ any probability measure on the set of path measures with values in $\mathbb{R}^n$ and satisfying $R_t = P_t$ for all $t \in [0,T]$. Then there holds*

$$H(P|W^\sigma) \leq H(R|W^\sigma),$$

with equality if and only if $P = R$.

This ensures that when the stochastic process generating the data is a SDE with a gradient drift, the Schrödinger problem (1.5) allows to find the measure of this SDE when the reference measure

is a Brownian motion (without drift), at least when the level of stochasticity is known and the number of observed cells is large enough. Thus, using the equality between the two problems (non-dynamical and dynamical), the authors were able to characterize an optimal drift w.r.t two timepoints, and then extended the method for an arbitrary number of timepoints. Knowing the characteristics of the SDE, they could therefore reconstruct the trajectory in the space of distributions. As detailed in Section 1.1.2, the optimal coupling solving the Schrödinger problem converges, when the level of noise goes to 0, to the solution of the optimal transport with a quadratic cost (which is the cost associated to the Brownian motion in the weak noise limit). The solution of such problem associates to any cell at a timepoint t_i an unique descendant cell at t_{i+1}, allowing to predict the most probable fate of a cell between two timepoints. Moreover, the most probable trajectory between each pair of cells is thus characterized by the deterministic system $\dot{X}_T = \nabla V(X_t)$, V being the potential of the SDE associated to the path measure solving the dynamical Schrödinger problem.
Importantly, this methodology has been extended (but with some heuristics) to the case with proliferation, when the master equation of the process generating the observations is supposed to be of the form (1.20) plus a proliferation term (as for the PBA method). To this end, the authors use a splitting scheme for solving at each timepoint alternatively the transport map and then a "proliferation map" explaining the change of total mass between datasets. To the best of our knowledge, this is the first dynamical method that takes into account stochasticity arising from transcriptional activity and proliferation at the same time. Note also that an alternative methodology has been proposed for the stationary case, when the dataset consists in one single snapshot, provided that a reliable estimation of the proliferation rates is available [107].
We will develop in Chapter 6 a framework similar to the one of Waddington-OT, using the connections between stochastic processes and Schrödinger problems, but using the crucial fact that cell dynamics is driven by a PDMP process of the form (1.16). Importantly, we will see that although very preliminary, our results suggest that such method is more suitable for assessing the relevance of mechanistic assumptions to data than for actually reconstructing cell dynamics or inferring process characteristics from minimal assumptions.

Some comments about methods with asymptotic guarantees

In all this section, we insisted on the importance of the number of observations that are used for reconstructing the features of a stochastic process. For some methods (PBA and Waddington-OT), we also mentioned the guarantee of asymptotic convergence, *i.e* the assurance that in case of an infinite number of observations, and under the additional hypotheses specific to the methods, the results are exact. Such asymptotic convergence is obviously considered as an important advantage for a given statistical method. It is nevertheless worth noticing that it does not systematically guarantees the efficiency of the procedure. For example, a method which uses information about low-probable events may have this kind of asymptotic guarantees, but if the number of observations is not high enough for estimating reliably these events (and it needs to be all the greater as these events are rare), the results will break down. Thus, if a method has no convergence bounds ensuring that the result that it provides is reliable when applied to a realistic number of observations, we believe that it is more important to evaluate it on its practical results than on its theoretical guarantees of asymptotic convergence. For example, as it will be mentioned in the following chapters, we believe that the methods developed in [73] and [104], which needs to estimate the potential function on areas of law probability of the gene expression space for reconstructing the transition between cell states, may be strongly inaccurate when we have not enough cells and/or that the transitions in the gene expression space are too rare.

Part I

Building an approximate landscape of cellular differentiation.

Chapter 2

Reduction of a mechanistic model of gene expression. Article published in Journal of Mathematical Biology.

The aim of this chapter is provide a mathematical framework able to link the behavior of a cell at the molecular level to the dynamics between observable cell types, in order to understand the global patters of cell differentiation as the emergent property of an underlying GRN. We will use the cell types characterization proposed by Huang [39], mentioned in the introduction, who considered that they correspond to the basins of attraction of the deterministic limit of the stochastic model describing differentiation. We are going to consider the PDMP model (1.15), and focus on proteins dynamics: the link with mRNA levels will be completed in the next chapter. For deriving a deterministic, we will use the experimentally-observed fact that promoters switches are fast regarding proteins dynamics. The scaling between these two dynamics plays the role of the factor scaling the diffusion coefficient in an SDE of the form (1.21). Indeed, considering the model (1.15), in the limit of infinitely fast promoters switches, proteins always behave as if promoters had reached their steady state (knowing proteins), as they reach it at the limit before proteins levels can change. This makes the mean behavior of proteins deterministic. Then, we will keep the classical notation ε for this scaling factor in analogy to Large deviations theory developed for SDEs [29].
We will first derive the deterministic limit for the rescaled PDMP model, and extend results from Large deviations theory to this model by showing that the rate function of a Large deviations principle shown in [25] can be found analytically from a spectral characterization developed in [14]. We will then use these results for developing a numerical method able to compute, from a given GRN, the cell types and the transition rates of the discrete Markov chain characterizing their dynamics. This reduction allow to deduce a new phenomenological model describing explicitly both the cell types and the cell dynamics in the gene expression space, which has the great advantage of having an explicit stationary distribution, allowing an approximation of the landscape. Altogether this work establishes a formal basis for the definition of an genetic/epigenetic landscape, given a GRN.

This chapter contains an article which has been published in *Journal of Mathematical biology* [97]. Note that we added a new appendix K to the published article that precises the link between the phenomenological model and the Beta-mixture that is used for approximating the landscape.

Journal of Mathematical Biology (2021) 83:59
https://doi.org/10.1007/s00285-021-01684-1

Mathematical Biology

Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastabilty

Elias Ventre[1,2,3] · Thibault Espinasse[2,3] · Charles-Edouard Bréhier[3] · Vincent Calvez[2,3] · Thomas Lepoutre[2,3] · Olivier Gandrillon[1,2]

Received: 21 September 2020 / Revised: 2 September 2021 / Accepted: 13 October 2021 /
Published online: 5 November 2021

Abstract

Differentiation is the process whereby a cell acquires a specific phenotype, by differential gene expression as a function of time. This is thought to result from the dynamical functioning of an underlying Gene Regulatory Network (GRN). The precise path from the stochastic GRN behavior to the resulting cell state is still an open question. In this work we propose to reduce a stochastic model of gene expression, where a cell is represented by a vector in a continuous space of gene expression, to a discrete coarse-grained model on a limited number of cell types. We develop analytical results and numerical tools to perform this reduction for a specific model characterizing the evolution of a cell by a system of piecewise deterministic Markov processes (PDMP). Solving a spectral problem, we find the explicit variational form of the rate function associated to a large deviations principle, for any number of genes. The resulting Lagrangian dynamics allows us to define a deterministic limit of which the basins of attraction can be identified to cellular types. In this context the quasipotential, describing the transitions between these basins in the weak noise limit, can be defined as the unique solution of an Hamilton–Jacobi equation under a particular constraint. We develop a numerical method for approximating the coarse-grained model parameters, and show its accuracy for a symmetric toggle-switch network. We deduce from the reduced model an approximation of the stationary distribution of the PDMP system, which appears as a Beta mixture. Altogether those results establish a rigorous

Elias Ventre
elias.ventre@ens-lyon.fr

1 ENS de Lyon, CNRS UMR 5239, Laboratory of Biology and Modelling of the Cell, Lyon, France

2 Inria Center Grenoble Rhone-Alpes, Team Dracula, Villeurbanne, France

3 Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, Villeurbanne, France

frame for connecting GRN behavior to the resulting cellular behavior, including the calculation of the probability of jumps between cell types.

Keywords Single cell · Gene regulation network · Energetic landscape · Piecewise deterministic Markov processes · Large deviations · Metastability

Mathematics Subject Classification 92C42 · 60J25 · 60F10

1 Introduction

Differentiation is the process whereby a cell acquires a specific phenotype, by differential gene expression as a function of time. Measuring how gene expression changes as differentiation proceeds is therefore of essence to understand this process. Advances in measurement technologies now allow to obtain gene expression levels at the single cell level. It offers a much more accurate view than population-based measurements, that has been obscured by mean population-based averaging (Mar 2019; Coskun et al. 2016). It has been established that there is a high cell-to-cell variability in gene expression, and that this variability has to be taken into account when investigating a differentiation process at the single-cell level (Moris and Arias 2017; Mohammed et al. 2017; Antolovic et al. 2017; Semrau et al. 2017; Mojtahedi et al. 2016; Richard et al. 2016; Moussy et al. 2017; Guillemin et al. 2019; Stumpf et al. 2017).

A popular vision of the cellular evolution during differentiation, introduced by Waddington in Waddington (1957), is to compare cells to marbles following probabilistic trajectories, as they roll through a developmental landscape of ridges and valleys. These trajectories are represented in the gene expression space: a cell can be described by a vector, each coordinate of which represents the expression of a gene (Huang and Ingber 2007; Moris et al. 2016). Thus, the state of a cell is characterized by its position in the gene expression space, i.e its specific level for all of its expressed genes. This landscape is often considered to be shaped by the underlying gene regulatory network (GRN), the behavior of which can be influenced by many factors, such as proliferation or cell-to-cell communication.

Theoretically, the number of states a cell can take is equal to the number of possible combination of protein quantities associated to each gene. This number is potentially huge (Braun 2015). But metastability seems inherent to cell differentiation processes, as evidenced by limited number of existing cellular phenotypes (Morris 2019; Bizzarri et al. 2018), providing a rationale for dimension reduction approaches (Moon et al. 2018). Indeed, since (Kauffman 2004) and (Huang et al. 2005), many authors have identified cell types with the basins of attraction of a dynamical system modeling the differentiation process, although the very concept of "cell type" has to be interrogated in the era of single-cell omics (Clevers et al. 2017).

Adapting this identification for characterizing metastability in the case of stochastic models of gene expression has been studied mostly in the context of stochastic diffusion processes (Wang et al. 2010, 2011; Zhou et al. 2012), but also for stochastic hybrid systems (Lin and Galla 2016). In the weak noise limit, a natural development of this

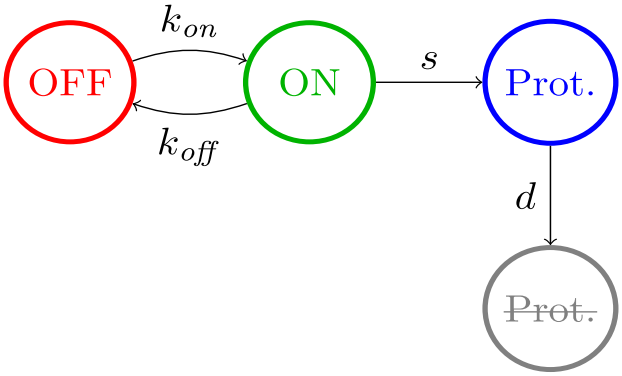


Fig. 1 Simplified two-states model of gene expression (Herbach et al. 2017; Peccoud and Ycart 1995) (colour figure online)

analysis consists in describing the transitions between different macrostates within the large deviations framework (Lv et al. 2014; Bressloff 2014).

We are going to apply this strategy for a piecewise-deterministic Markov process (PDMP) describing GRN dynamics within a single cell, introduced in Herbach et al. (2017), which corresponds accurately to the non-Gaussian distribution of single-cell gene expression data. Using the work of Bressloff and Faugeras (2017), the novelty of this article is to provide analytical results for characterizing the metastable behavior of the model for any number of genes, and to combine them with a numerical analysis for performing the reduction of the model in a coarse-grained discrete process on cell types. We detail the model in Sect. 2, and we present in Sect. 3 how the reduction of this model in a continuous-time Markov chain on cell types allows to characterize the notion of metastability. For an arbitrary network, we provide in Sect. 4.1 a numerical method for approximating each transition rate of this coarse-grained model, depending on the probability of a rare event. In Sect. 4.2, we show that this probability is linked to a large deviations principle. The main contribution of this article is to derive in Sect. 5.1 the explicit variational form of the rate function associated to a Large deviations principle (LDP) for this model. We discuss in Secs. 5.2 and 5.3 the conditions for which a unique quasipotential exists and allows to describe transitions between basins. We replace in Sect. 5.4 these results in the context of studying metastability. Finally, we apply in Sect. 6 the general results to a toggle-switch network. We also discuss in Sect. 7.1 some notions of energy associated to the LDP and we propose in Sect. 7.2 a non-Gaussian mixture model for approximating proteins distribution.

2 Model description

The model which is used throughout this article is based on a hybrid version of the well-established two-state model of gene expression (Ko 1991; Peccoud and Ycart 1995). A gene is described by the state of the promoter, which can be $\{on, off\}$. If the state of the promoter is *on*, mRNAs are transcripted and translated into proteins, which are considered to be produced at a rate s. If the state of the promoter is *off*, only degradation of proteins occurs at a rate d (see Fig. 1). k_{on} and k_{off} denote the exponential rates of transition between the states *on* and *off*. This model is a reduction of a mechanistic model including both mRNA and proteins, which is described in "Appendix A".

Neglecting the molecular noise associated to proteins quantity, we obtain the hybrid model:

$$\begin{cases} E(t) : 0 \xrightarrow{k_{on}} 1, 1 \xrightarrow{k_{off}} 0, \\ P'(t) = sE(t) - dP(t). \end{cases}$$

where $E(t)$ denotes the promoter state, $P(t)$ denotes the protein concentration at time t, and we identify the state *off* with 0, the state *on* with 1.

The key idea for studying a GRN is to embed this two-states model into a network. Denoting the number of genes by n, the vector (E, P) describing the process is then of dimension $2n$. The jump rates for each gene i are expressed in terms of two specific functions $k_{on,i}$ and $k_{off,i}$. To take into account the interactions between the genes, we consider that for all $i = 1, \ldots, n$, $k_{on,i}$ is a function which depends on the full vector P via the GRN, represented by a matrix Θ of size n. We assume that $k_{on,i}$ is upper and lower bounded by a positive constant for all i. The function is chosen such that if gene i activates gene j, then $\partial_{P_i} k_{on,j} \geq 0$. For the sake of simplicity, we consider that $k_{off,i}$ does not depend on the protein level.

We introduce a typical time scale $\bar{k}$ for the rates of promoters activation $k_{on,i}$, and a typical time scale $\bar{d}$ for the rates of proteins degradation. Then, we define the scaling factor $\varepsilon = \frac{\bar{d}}{\bar{k}}$ which characterizes the difference in dynamics between two processes: 1. gene bursting dynamics and 2. protein dynamics. It is generally considered that promoter switches are fast with respect to protein dynamics, i.e that $\varepsilon \ll 1$, at least for eukaryotes (Suter et al. 2011). Driven by biological considerations, we will consider values of ε smaller than $1/5$ (see "Appendix A").

We then rescale the time of the process by $\bar{d}$. We also rescale the quantities $k_{on,i}$ and $k_{off,i}$ by $\bar{k}$, and d_i by $\bar{d}$, for any gene i, in order to simplify the notations. Finally, the parameters s_i can be removed by a simple rescaling of the protein concentration P_i for every gene by its equilibrium value when $E_i = 1$ (see Herbach et al. 2017 for more details). We obtain a reduced dimensionless PDMP system modeling the expression of n genes in a single cell:

$$\forall i = 1, \ldots, n : \begin{cases} E_i(t) : 0 \xrightarrow{\frac{k_{on,i}(X(t))}{\varepsilon}} 1, 1 \xrightarrow{\frac{k_{off,i}}{\varepsilon}} 0, \\ X_i'(t) = d_i(E_i(t) - X_i(t)). \end{cases} \tag{1}$$

Here, X describes the protein vector in the renormalized gene expression space $\Omega := (0, 1)^n$ and E describes the promoters state, in $P_E := \{0, 1\}^n$. We will refer to this model, that we will use throughout this article, as the PDMP system.

As $\text{card}(P_E) = 2^n$, we can write the joint probability density $u(t, e, x)$ of (E_t, X_t) as a 2^n-dimensional vector $u(t, x) = (u_e(t, x))_{e \in P_E} \in \mathbb{R}^{2^n}$. The master equation on u can be written:

$$\frac{\partial u}{\partial t}(t, x) + \sum_{i=1}^{n} \frac{\partial}{\partial x_i} (F_i(x) u(t, x)) = \frac{1}{\varepsilon} \sum_{i=1}^{n} K_i(x) u(t, x). \tag{2}$$

For all $i = 1, \ldots, n$, for all $x \in \Omega$, $F_i(x)$ and $K_i(x)$ are matrices of size 2^n. Each F_i is diagonal, and the term on a line associated to a promoter state e corresponds to the drift of gene i: $d_i(e_i - x_i)$. K_i is not diagonal: each state e is coupled with every state e' such that only the coordinate e_i changes in e, from 1 to 0 or conversely. Each of these matrices can be expressed as a tensorial product of $(n-1)$ two-dimensional identity matrices with a two-dimensional matrix corresponding to the operator associated to an isolated gene:

- $F_i(x) = I_2 \otimes \cdots \otimes \underbrace{F^{(i)}(x)}_{i^{th}\text{ position}} \otimes \cdots \otimes I_2$
- $F^{(i)}(x) = \begin{pmatrix} -d_i x_i & 0 \\ 0 & d_i(1-x_i) \end{pmatrix}$
- $K_i(x) = I_2 \otimes \cdots \otimes \underbrace{K^{(i)}(x)}_{i^{th}\text{ position}} \otimes \cdots \otimes I_2,$
- $K^{(i)}(x) = \begin{pmatrix} -k_{on,i}(x) & k_{off,i}(x) \\ k_{on,i}(x) & -k_{off,i}(x) \end{pmatrix}.$

We detail in Appendix B the case of $n = 2$ for a better understanding of this tensorial expression.

3 Model reduction in the small noise limit

3.1 Deterministic approximation

The model (1) describes the promoter state of every gene i at every time as a Bernoulli random variable. We use the biological fact that promoter switches are frequent compared to protein dynamic, i.e $\varepsilon < 1$ with the previous notations. When $\varepsilon \ll 1$, we can approximate the conditional distribution of the promoters knowing proteins, ρ, by its quasistationary approximation $\overline{\rho}$:

$$\forall i = 1, \ldots, n,\ \forall x \in \Omega :\ \rho_i(x) \simeq \overline{\rho}_i(x) = \frac{k_{on,i}(x)}{k_{off,i} + k_{on,i}(x)}, \tag{3}$$

which is derived from the stationary distribution of the Markov chain on the promoters states, defined for a given value of the protein vector $X = x$ by the matrix $\sum_{i=1}^{n} K_i(x)$ (see Papanicolaou 1975; Newby and Keener 2011).

Thus, the PDMP model (1) can be coarsely approximated by a system of ordinary differential equations:

$$\forall i = 1, \ldots, n :\ \dot{x}_i(t) = d_i \left(\frac{k_{on,i}(x(t))}{k_{off,i} + k_{on,i}(x(t))} - x_i(t) \right). \tag{4}$$

Intuitively, these trajectories correspond to the mean behaviour of a cell in the weak noise limit, i.e when promoters jump much faster than proteins concentration changes. More precisely, a random path X_t^{ε} converges in probability to a trajectory ϕ_t solution of the system (4), when $\varepsilon \to 0$ (Faggionato et al. 2009). The diffusion limit, which keeps a residual noise scaled by $\sqrt{\varepsilon}$, can also be rigorously derived from the PDMP system (Pakdaman et al. 2012), which is detailed in "Appendix C.1".

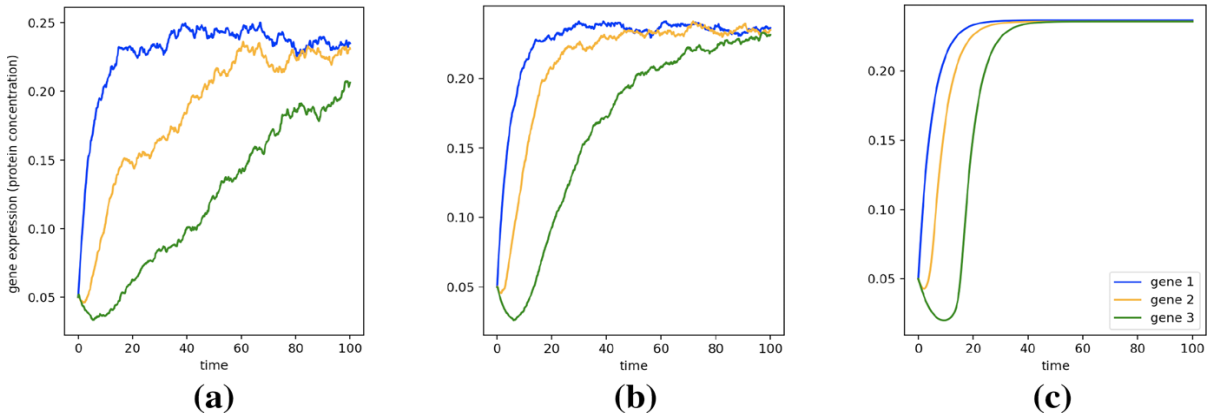


Fig. 2 Comparison between the average on 100 simulated trajectories with $\varepsilon = 1/7$ (**a**), $\varepsilon = 1/30$ (**b**) and the trajectories generated by the deterministic system (**c**) for a single pathway network: gene 1 $\longrightarrow$ gene 2 $\longrightarrow$ gene 3

In the sequel, we assume that every limit set of a trajectory solution of the system (4) as $t \to +\infty$ is reduced to a single equilibrium point, described by one of the solutions of:

$$\forall i = 1, \ldots, n : \frac{k_{on,i}(x)}{k_{off,i} + k_{on,i}(x)} - x_i = 0. \tag{5}$$

Note that the condition above strongly depends on the interaction functions $\{k_{on,i}\}_{i=1,\ldots,n}$. Alternatively speaking, in this work we rule out the existence of attractive limit cycles or more complicated orbits. We also assume that the closure of the basins of attraction which are associated to the stable equilibria of the system (5) covers the gene expression space Ω.

Without noise, the fate of a cell trajectory is fully characterized by its initial state x_0. Generically, it converges to the attractor of the basin of attraction it belongs to, which is a single point by assumption. However, noise can modify the deterministic trajectories in at least two ways. First, in short times, a stochastic trajectory can deviate significantly from the deterministic one. In the case of a single, global, attractor, the deterministic system generally allows to retrieve the global dynamics of the process, i.e the equilibrium and the order of convergence between the different genes, for realistic ε (see Fig. 2).

Second, in long times, stochastic dynamics can even push the trajectory out of the basin of attraction of one equilibrium state to another one, changing radically the fate of the cell. These transitions cannot be catched by the deterministic limit, and happen on a time scale which is expected to be of the order of $e^{\frac{C}{\varepsilon}}$ (owing to a Large deviations principle studied below), where C is an unknown constant depending on the basins. In Fig. 3a, we illustrate this situation for a toggle-switch network of two genes. We observe possible transitions between three basins of attraction. Two examples of random paths, the stochastic evolution of promoters and proteins along time, are represented in Fig. 3b, c for different values of ε. All the details on the interaction functions and the parameters used for this network can be found respectively in the "Appendices D and E".

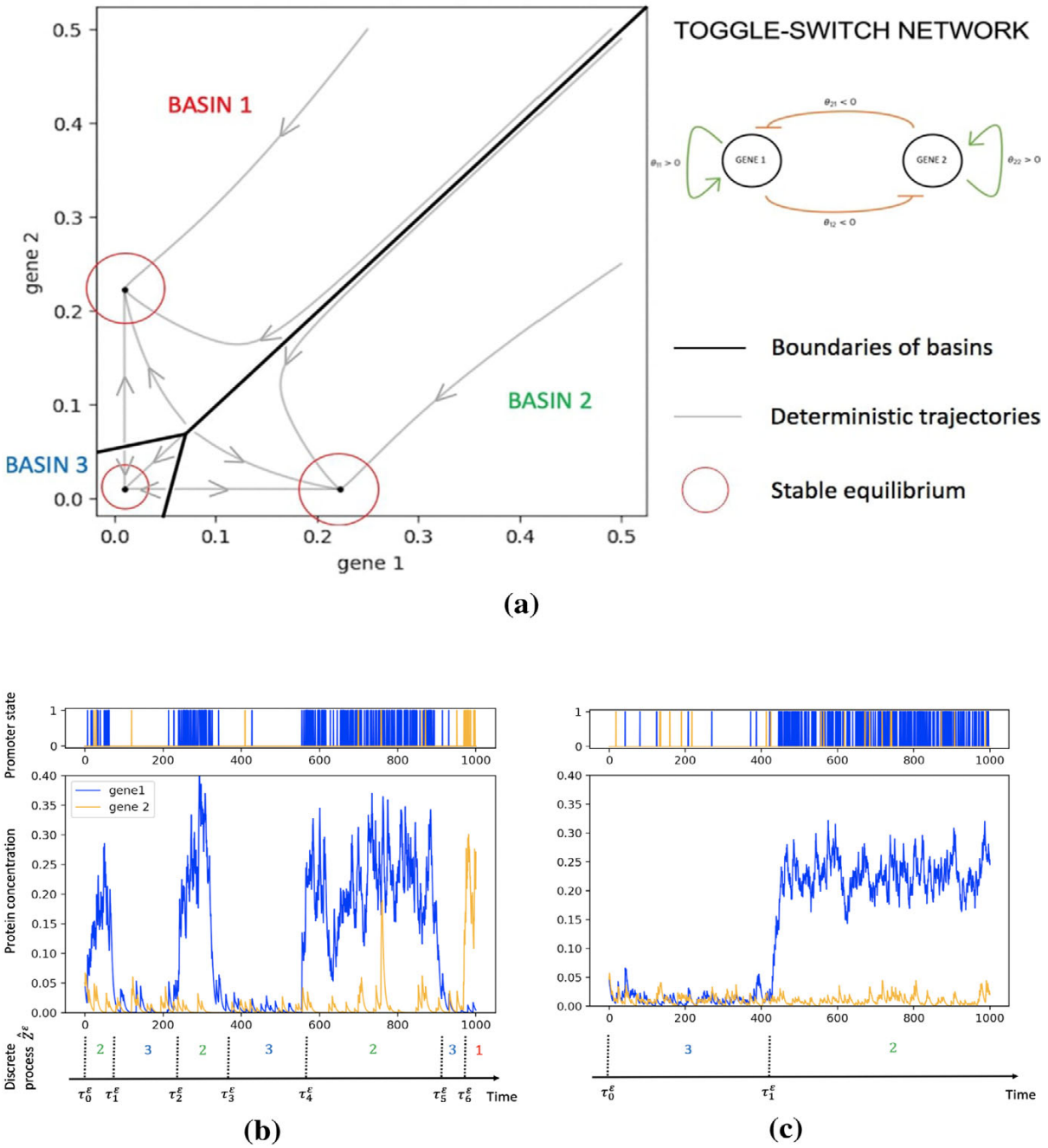


Fig. 3 **a** Phase portrait of the deterministic approximation for a symmetric toggle-switch with strong inhibition: two genes which activate themselves and inhibit each other. **b** Example of a stochastic trajectory generated by the toggle-switch, for $\varepsilon = 1/7$. **c** Example of a stochastic trajectory generated by the toggle-switch, for $\varepsilon = 1/30$

3.2 Metastability

When the parameter ε is small, transitions from one basin of attraction to another are rare events: in fact the mean escape time from each basin is much larger than the time required to reach a local equilibrium (quasi-stationary state) in the basin.

Adopting the paradigm of metastability mentioned in the introduction, we identify each cell type to a basin of attraction associated to a stable equilibrium of the determin-

istic system (4). In this point of view, a cell type corresponds to a metastable sub-region of the gene expression space. It also corresponds to the notion of macrostate used in the theory of Markov State Models which has been recently applied to a discrete cell differentiation model in Chu et al. (2017). Provided we can describe accurately the rates of transition between the basins on the long run, the process can then be coarsely reduced to a new discrete process on the cell types.

More precisely, let m be the number of stable equilibrium points of the system (4), called attractors. We denote Z the set of the m basins of attraction associated to these m attractors, that we order arbitrarily: $Z = \{Z_1, \cdots, Z_m\}$. The attractors are denoted by $(X_{eq,Z_i})_{Z_i \in Z}$. Each attractor is associated to a unique basin of attraction. By assumption, the closure of these m basins, $\overline{Z} = \{\overline{Z_1}, \cdots, \overline{Z_m}\}$, covers the gene expression space Ω. To obtain an explicit characterization of the metastable behavior, we are going to build a discrete process $\hat{Z}^\varepsilon$, with values in Z. From a random path X_t^ε of the PDMP system such that $X_0^\varepsilon \in Z$, we define a discrete process $\hat{Z}^\varepsilon$ describing the cell types:

$$\forall l \in \mathbb{N} : \ \hat{Z}_l^\varepsilon = \sum_{i=1}^{m} Z_i \mathbb{1}_{\{X_{\tau_l}^\varepsilon \in Z_i\}},$$

where $(\tau_l^\varepsilon)_{l \in \mathbb{N}}$ is a sequence of stopping times defined by:
$\tau_0^\varepsilon = 0, \forall l \in \mathbb{N}^* : \ \tau_l^\varepsilon = \inf\{t \geq \tau_{l-1}^\varepsilon \mid X_t^\varepsilon \in Z \setminus \hat{Z}_{l-1}^\varepsilon\}$. Note that $\hat{Z}_l^\varepsilon$ are the successive metastable states, and that τ_l^ε are the successive times of transition between them. From the convergence of any random path to a solution of the deterministic system (4) , that we mentioned in Sect. 3.1, we know that for every basin Z_i such that $\hat{Z}_l = Z_i$, whatever is the point on the boundary of Z_i which has been first attained, X_t^ε reaches any small neighborhood of the attractor X_{eq,Z_i} of Z_i before leaving the basin, with probability converging to 1 as $\varepsilon \to 0$. In addition, for any basin Z_j, the probability $\mathbb{P}(\hat{Z}_{l+1} = Z_j)$ is asymptotically independent of the value of $X_{\tau_l}^\varepsilon$, and is then asymptotically determined by the value of $\hat{Z}_l$. In other words, $\hat{Z}^\varepsilon$ converges to a Markov chain when $\varepsilon \to 0$. We refer to Kurtz and Swanson (2019) to go further in the analysis of the coupling between the processes $\hat{Z}^\varepsilon$ and X^ε for general Markov processes.

For small ε, it is natural to approximate the distribution of the exit time from a basin by an exponential distribution. The *in silico* distribution represented in Fig. 4 suggests that this assumption seems accurate for the toggle-switch network, even for a realistic value of ε. Note, however, that the exponential approximation slightly overestimates the probability that the exit times are small.

To completely characterize the coarse-grained resulting process, it remains to compute the transition rates $\{a_{ij}^\varepsilon\}_{i,j}$ of the time-continuous Markov chain on the basins, that we define for all pair of basins $(Z_i, Z_j) \in Z^2$, $i \neq j$, by:

$$a_{ij}^\varepsilon = \frac{\mathbb{P}(\hat{Z}_1^\varepsilon = Z_j \mid \hat{Z}_0^\varepsilon = Z_i)}{\mathbb{E}(\tau_1^\varepsilon \mid \hat{Z}_0^\varepsilon = Z_i)}, \tag{6}$$

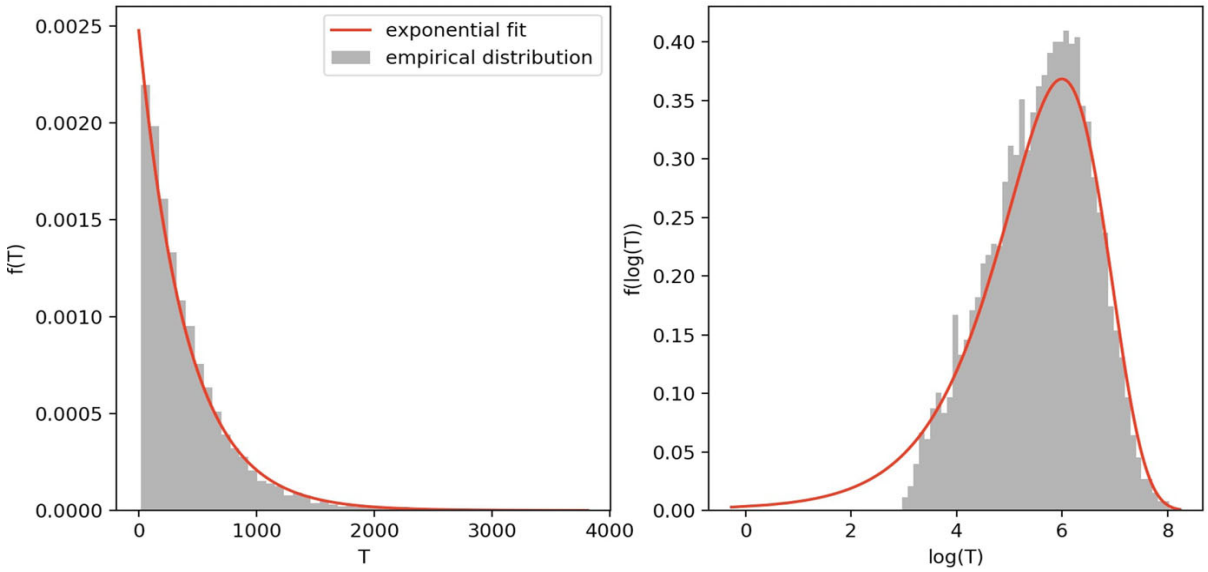


Fig. 4 Comparison between the distribution of the exit time from a basin, obtained with a Monte-Carlo method, and the exponential law with appropriate expected value, for $\varepsilon = 1/7$. We represent the two densities in normal scale (on the left-hand side) and in logarithmic scale (on the right-hand side) to observe that the exponential law overestimates the probability that the exit times are small

where $\mathbb{E}(\tau_1^\varepsilon \mid \hat{Z}_0^\varepsilon = Z_i)$ is called the Mean First Exit Time of exit from Z_i. This Markov process with discrete state space Z represents accurately, when ε is small enough, the main dynamics of the metastable system in the weak noise limit (Freidlin and Wentzell 2012). This reduced model is fully described by m^2 transition rates: when the number of genes n is large, it is significantly smaller than the n^2 parameters characterizing the GRN model (see "Appendix D").

This collection of transition rates are characterized by rare events: when $\varepsilon \ll 1$ or when the number of genes is large, it would be too expensive to compute them with a crude Monte-Carlo method. We are then going to present a method for approximating these transition rates from probabilities of some rare events. We will detail afterwards how these probabilities can be computed either by an efficient numerical method or an analytical approximation.

4 Computing the transition rates

4.1 Transition rates from probabilities of rare events

In this Section, we approximate each transition rate between any pair of basins (Z_i, Z_j), $j \neq i$ in terms of the probability that a random path realizes a certain rare event in the weak noise limit.

Let us consider two small parameters r, R such that $0 < r < R$. We denote γ_{Z_i} the r-neighborhood of X_{eq,Z_i}, and Γ_{Z_i} its R-neighborhood. For a random path X_t^ε of the PDMP system starting in $x_0 \in \partial\Gamma_{Z_i}$, we denote the probability of reaching a basin Z_j, $j \neq i$ before any other basin Z_k, $k \neq i$, j, and before entering in γ_{Z_i}:

$$p_{ij}^{\varepsilon}(x_0) = \mathbb{P}_{x_0}\left(T_{Z_j}^{\varepsilon} < T_{\gamma_{Z_i}\cup\{Z\setminus\{\overline{Z_i}\cup\overline{Z_j}\}\}}^{\varepsilon}\right), \tag{7}$$

where $T_A^{\varepsilon} = \inf\{t \geq 0 \mid X_t^{\varepsilon} \in A\}$ is the hitting time of a set $A \subset \Omega$.

The method developed in Cérou et al. (2011) aims to show how it is possible, from the knowledge of the probability (7), to approximate the transition rate a_{ij}^{ε} presented in (6). Briefly, it consists in cutting a transition path into two pieces, a piece going from X_{eq,Z_i} to $\partial\Gamma_{Z_i}$ and another reaching Z_j from $\partial\Gamma_{Z_i}$: the transition rates a_{ij} can be then approximated by the reverse of the mean number of attempts of reaching Z_j from Γ_{Z_i} before entering in γ_{Z_i}, which is close to the inverse of the rare event probability given by (7) when $x_0 \in \partial\Gamma_{Z_i}$, multiplied by the average time of each excursion, that we denote $\overline{T}_{Z_i,Z_i}^{\varepsilon}$. We obtain:

$$a_{ij}^{\varepsilon} \simeq \frac{p_{ij}^{\varepsilon}(x_0)}{\overline{T}_{Z_i,Z_i}^{\varepsilon}}. \tag{8}$$

It is worth noticing that to be rigorous, this method need to redefine the neighborhoods γ_{Z_i} and Γ_{Z_i} by substituting to the squared euclidean distance a new function based on the probability of reaching the (unknown) boundary: $\forall x, y \in Z_i, \; || x - y ||^2 \leftarrow | p_{ij}^{\varepsilon}(x) - p_{ij}^{\varepsilon}(y) |$. The details are provided in "Appendix F".

We observe that the average time $\overline{T}_{Z_i,Z_i}^{\varepsilon}$ can be easily computed by a crude Monte-Carlo method: indeed, the trajectories entering in γ_{Z_i} are not rare. It thus only remains to explain the last ingredient for the approximation of the transition rates, which is how to estimate the probabilities of the form (7).

4.2 Computing probabilities of the form (7)

A powerful method for computing probabilities of rare events like (7), is given by splitting algorithms. We decide to adapt the Adaptative Multilevel Splitting Algorithm (AMS) described in Bréhier et al. (2016) to the PDMP system: all the details concerning this algorithm can be found in Appendix G. In Sect. 6, we will verify that the probabilities given by the AMS algorithm are consistent with the ones obtained by a crude Monte-Carlo method for the toggle-switch network.

However, estimating the probability (7) becomes hard when both the number of genes of interest increases and ε decreases. Indeed, the AMS algorithm allows to compute probabilities much smaller than the ones we expect for biologically relevant parameters ($\varepsilon \approx 0.1$), but the needed number of samples grows at least with a polynomial factor in ε^{-1}. If the number of genes considered is large, these simulations can make the algorithm impossible to run in a reasonable time. A precise analysis of the scalability of this method for the PDMP system is beyond the scope of this article, but we have been able to get consistent results on a laptop in time less than one hour for a network of 5 genes, with $\varepsilon > 1/15$. The resulting probabilities were of order 5.10^{-3}.

In order to overcome this problem, we are now going to develop an analytical approach for approximating these probabilities, by means of optimal trajectories exiting each basin of attraction. The later can be computed within the context of Large

deviations. As we will see in Sect. 6, this approach is complementary to the AMS algorithm, and consistent with it.

4.2.1 Large deviations setting

In this section, we derive a variational principle for approximating the transition probability (7) introduced in Sect. 4.1. A powerful methodology for rigorously deriving a variational principle for optimal paths is Large deviations theory. It has been developed extensively within the context of Stochastic Differential Equations (SDE) (Freidlin and Wentzell 2012; Dembod et al. 1996). For the sake of simplicity, we present here only an heuristic version. There exists a Large deviations principle (LDP) for a stochastic process with value in Ω if for all $x_0 \in \Omega$ there exists a lower semi-continuous function defined on the set of continuous trajectories from $[0, T]$ to Ω, $J_T : C_{0T}(\mathbb{R}^n) \to [0, \infty]$, such that for all set of trajectories $A \subset C_{0T}(\mathbb{R}^n)$:

$$-\varepsilon \ln \left(\mathbb{P}^{\varepsilon}_{x_0}(X^{\varepsilon}_t \in A) \right) \underset{\varepsilon \to 0}{\to} \min_{\phi \in A} J_T(\phi). \tag{9}$$

The function J_T is called the rate function of the process in $[0, T]$, and the quantity $J_T(\phi)$ is called the cost of the trajectory ϕ over $[0, T]$.

The particular application of this theory to stochastic hybrid systems has been developed in detail in Kifer (2009) and Faggionato et al. (2009). We now present consequences of results developed in Bressloff and Faugeras (2017).

Definition 1 The Hamiltonian is the function $H : \Omega \times \mathbb{R}^n \mapsto \mathbb{R}$, such that for all $(x, p) \in \Omega \times \mathbb{R}^n$, $H(x, p)$ is the unique eigenvalue associated to a nonnegative right-eigenvector $\zeta(x, p) \in \mathbb{R}^{2^n}$, (which is unique up to a normalization), of the following spectral problem:

$$M(x, p)\zeta(x, p) = H(x, p)\zeta(x, p), \tag{10}$$

where the matrix $M(x, p) \in M_{2^n, 2^n}(\mathbb{R})$ is defined by:

$$M(x, p) = \sum_{i=1}^{n} \left(K_i(x) + p_i F_i(x) \right). \tag{11}$$

We remark that the matrix $M(x, p)$ has off-diagonal nonnegative coefficients. Moreover, the positivity of the functions $k_{on,i}$ makes M irreducible (the matrix allows transition between any pair $(e, e') \in P_E^2$ after at most n steps). Thereby, the Perron Frobenius Theorem may be applied, and it justifies the existence and uniqueness of $H(x, p)$. Moreover, from a general property of Perron eigenvalues when the variable p appears only on the diagonal of the matrix, H is known to be convex (Cohen 1981).

The following result is a direct consequence of theoretical results of Bressloff and Faugeras (2017) applied to the PDMP system (1).

Theorem 1 *Let us denote* $\Omega^v(x) = \bigotimes_{i=1}^{n} [-d_i x_i, d_i(1 - x_i)]$ *and* $\mathring{\Omega}^v(x)$ *its interior.*

The Fenchel-Legendre transform of the Hamiltonian H is well-defined and satisfies: $\forall x \in \Omega$, $\forall v \in \mathring{\Omega}^v(x)$,

$$L(x, v) = \sup_{p \in \mathbb{R}^n} \left(\langle p, v \rangle - H(x, p) \right).$$

Moreover, the PDMP system (1) *satisfies a LDP, and the associated rate function has the form of a classical action: the cost of any piecewise differentiable trajectory ϕ_t in $C_{0T}(\mathbb{R}^n)$ satisfying for all $t \in [0, T)$, $\dot{\phi}(t) \in \mathring{\Omega}^v(\phi(t))$, can be expressed as*

$$J_T(\phi) = \int_0^T L(\phi(t), \dot{\phi}(t)) dt. \tag{12}$$

The function L is called the Lagrangian of the PDMP system.

We are now going to show how in certain cases, trajectories which minimize the quantity (12) between two sets in Ω can be defined with the help of solutions of an Hamilton–Jacobi equation with Hamiltonian H.

4.2.2 WKB approximation and Hamilton–Jacobi equation

The Hamiltonian defined in (10) also appears in the WKB (Wentzell, Kramer, Brillouin) approximation of the master equation (Newby and Keener 2011; Bressloff and Faugeras 2017). This approximation consists in injecting in the master equation (2) of the PDMP system, a solution of the form:

$$\forall e \in P_E,\ u_e(x, t) = \pi_e(x, t) e^{-\frac{S(x,t)}{\varepsilon}}, \tag{13}$$

where $e^{-\frac{S(\cdot,t)}{\varepsilon}}$ then denotes the marginal distribution on proteins of the distribution u at time t, and $\pi_e(x, t)$ is a probability vector denoting the conditional distribution of the promoters knowing that proteins are fixed to $X = x$ at t. The expression (13) is justified under the assumption that the density u_e is positive at all times.

Under the regularity assumptions $S \in C^1(\Omega \times \mathbb{R}^+, \mathbb{R})$ and $\pi \in \left(C^1(\Omega \times \mathbb{R}^+, \mathbb{R})\right)^{2^n}$, we can perform a Taylor expansion in ε of the functions S and π_e, for any $e \in P_E$, and keeping in the resulting master equation only the leading order terms in ε, that we denote S_0 and $\pi_0 = (\pi_{0,e})_{e \in P_E}$, we obtain:

$$-\partial_t S_0(x, t) \pi_0(x, t) = \sum_{i=1}^{n} \left(K_i(x) + \partial_{x_i}(S_0) F_i(x) \right) \pi_0(x, t).$$

Identifying the vectors π_0 and $\nabla_x S_0$ with the variables ζ and p in Eq. (10), we obtain that S_0 is solution of an Hamilton–Jacobi equation:

$$\forall x \in \Omega :\ H(x, \nabla_x S_0(x, t)) + \partial_t S_0(x, t) = 0. \tag{14}$$

More precisely, if at any time t, the marginal distribution on proteins of the PDMP process, denoted $u(\cdot, t)$, follows a LDP and if its rate function, denoted $V(\cdot, t)$, is differentiable on Ω, then the function $H(\cdot, \nabla_x V(\cdot, t))$ appears as the time derivative of $V(\cdot, t)$ at t. Moreover, the WKB method presented above shows that the rate function $V(\cdot, t)$ is identified for any time t with the leading order approximation in ε of the function $S(\cdot, t) = -\varepsilon \log(u(\cdot, t))$. Note that (13) is also reminiscent of the Gibbs distribution associated with a potential S. Some details about the interpretation of this equation and its link with the quasistationary approximation can be found in Newby and Keener (2011).

Next, we consider a function V, solution of the Hamilton–Jacobi Eq. (14), that we assume being of class $C^1(\Omega \times \mathbb{R}^+, \mathbb{R})$ for the sake of simplicity. Then, for any piecewise differentiable trajectory $\phi_t \in C_{0T}(\Omega)$ such that $\dot{\phi}(t) \in \Omega^v(\phi(t))$ for all $t \in [0, T)$, one has, by definition of the Fenchel-Legendre transform:

$$\begin{aligned}
\int_0^T L(\phi(t), \dot{\phi}(t))dt &= \int_0^T \sup_p \left(\sum_{i=1}^n p_i \dot{\phi}_i(t) - H(\phi(t), p) \right) dt \\
&\geq \int_0^T \left(\sum_{i=1}^n \partial_{x_i} V(\phi(t), t)\dot{\phi}_i(t) - H(\phi(t), \nabla_x V(\phi(t), t)) \right) dt \\
&= \int_0^T \left(\sum_{i=1}^n \partial_{x_i} V(\phi(t), t)\dot{\phi}_i(t) + \partial_t V(\phi(t), t) \right) dt \\
&= V(\phi(T), T) - V(\phi(0), 0). \qquad (15)
\end{aligned}$$

Moreover, when H is strictly convex in p, we have:

$$\forall v \in \Omega^v(x), \forall x, p \in \mathbb{R}^n \times \mathbb{R}^n, (L(x, v) = \langle p, v \rangle - H(x, p)) \iff \left(v = \nabla_p H(x, p)\right).$$

Then, the equality in (15) is exactly reached at any time for trajectories ϕ_t such that for all $t \in [0, T)$, $i = 1, \ldots, n$:

$$\begin{cases} p_i(t) &= \partial_{x_i} V(\phi(t), t), \\ \dot{\phi}_i(t) &= \partial_{p_i} H(\phi(t), p(t)). \end{cases} \qquad (16)$$

4.2.3 General method for computing probabilities of the form (7)

We now detail the link existing between the regular solutions V of the Hamilton–Jacobi Eq. (14) and the probabilities of the form (7). For this, we introduce the notion of quasipotential.

Definition 2 Denoting $C_{0T}^{1,pw}(\Omega)$ the set of piecewise differentiable trajectories in $C_{0T}(\Omega)$, we define the quasipotential as follows: for two sets $A, B \subset \Omega$ and a set $R \subset \Omega \setminus (A \cup B)$,

$$Q_R(A, B) = \inf_{\phi_t, T} \{J_T(\phi) \mid \phi_t \in C_{0T}^{1,pw}(\Omega), \phi(0) \in A, \phi(T) \in B, \forall t \in (0, T) : \phi(t) \notin R\}.$$

We call a trajectory $\phi_t \in C_{0T}^{1,pw}(\Omega)$ an optimal trajectory between the two subsets $A, B \subset \Omega$ in $\Omega \setminus R$, if it reaches the previous infimum.

For the sake of simplicity, if $R = \varnothing$, we will write $Q_R(A, B) = Q(A, B)$.

With these notations, the LDP principle allows to approximate for any basin Z_j, $i \neq j$, the probability $p_{ij}^{\varepsilon}(x_0)$ defined in (7), which is the probability of reaching Z_j, from a point $x_0 \in \Gamma_{min,Z_i}$, before γ_{min,Z_i}, by the expression:

$$-\varepsilon \ln\left(p_{ij}^{\varepsilon}(x_0)\right) \underset{\varepsilon \to 0}{\to} Q_{\bigcup\limits_{k \neq i,j} \{\partial Z_i \cap \partial Z_k\}}(x_0, \partial Z_i \cap \partial Z_j).$$

A direct consequence of the inequality (15), in the case where equality is reached, is that a regular solution V of the Hamilton–Jacobi Eq. (14) defines trajectories for which the cost (12) is minimal between any pair of its points. Moreover, if V is a stationary solution of (14), the cost of such trajectories does not depend on time: these trajectories are then optimal between any pair of its points among every trajectory in any time. We immediately deduce the following lemma:

Lemma 1 *For a stationary solution $V \in C^1(\Omega, \mathbb{R})$ of (14) and for all $T > 0$, any trajectory $\phi_t \in C_{0T}^{1,pw}(\Omega)$ satisfying the system (16) associated to V is optimal in Ω between $\phi(0)$ and $\phi(T)$, and we have:*

$$Q(\phi(0), \phi(T)) = J_T(\phi) = V(\phi(T)) - V(\phi(0)).$$

Thus, for approximating the probability of interest (7), between any pair of basin (Z_i, Z_j), we are going to build a trajectory Φ_t^{ij}, which verifies the system (16) associated to a stationary solution V of (14), with $\Phi^{ij}(0) = x_0 \in \Gamma_{min,Z_i}$, and which reaches in a time T a point $x \in \partial Z_i \cap \partial Z_j$ such that $Q(x_0, x) = Q_{\bigcup\limits_{k \neq i,j} \{\partial Z_i \cap \partial Z_k\}}(x_0, \partial Z_i \cap \partial Z_j)$. For such trajectory, from Lemma 1, we could then approximate the probability (7) by the formula

$$p_{ij}^{\varepsilon}(x_0) \simeq C_{ij} e^{\frac{-J_T(\Phi_t^{ij})}{\varepsilon}}, \tag{17}$$

where C_{ij} is an appropriate prefactor. Unfortunately, if there exists an explicit expression of C_{ij} in the one-dimensional case (Newby and Keener 2011), and that an approximation has been built for multi-dimensional SDE model (Bouchet and Reygner 2016), they are intractable or not applicable in our case. In general, the prefactor does not depend on ε (Berglund 2011). In that case $-\ln\left(p_{ij}^{\varepsilon}(x_0)\right)$ is asymptotically an affine function of ε^{-1}, the slope of which is $J_T(\Phi_t^{ij})$ and the initial value $-\ln(C_{ij})$. Then, the strategy we propose simply consists in approximating the prefactor by comparison between the probabilities given by the AMS algorithm and the Large deviations approximation (17) for a fixed ε (large enough to be numerically computed.)

To conclude, for every pair of basins (Z_i, Z_j), $i \neq j$, one of the most efficient methods for computing the probability (7) is to use the AMS algorithm. When the

dimension is large, and for values of ε which are too small for this algorithm to be efficiently run, we can use the LDP approximation (17), provided that the corresponding optimal trajectories Φ_t^{ij} can be explicitly found. The latter condition is studied in the next sections. The AMS algorithm is then still employed to approximate the prefactor, which is done using intermediate values of ε by the regression procedure mentioned above.

5 Analytical approximation of probabilities of the form (7) for the PDMP system

5.1 Expressions of the Hamiltonian and the Lagrangian

In this section, we identify the Perron eigenvalue $H(x, p)$ of the spectral problem (10), and prove that its Fenchel-Legendre transform L with respect to the variable p is well defined on $\mathbb{R}^n$. We then obtain the explicit form of the Hamiltonian and the Lagrangian associated to the LDP for the PDMP system (1).

Theorem 2 *For all n in $\mathbb{N}^*$, the Hamiltonian is expressed as follows: for all $(x, p) \in \Omega \times \mathbb{R}^n$, the unique solution of the spectral problem (10) (with nonnegative eigenvector) is:*

$$H(x, p) = \frac{1}{2} \sum_{i=1}^{n} \Big(p_i d_i (1 - 2x_i) - (k_{on,i}(x) + k_{off,i}) + \sqrt{(p_i d_i + k_{on,i}(x) - k_{off,i})^2 + 4 k_{on,i}(x) k_{off,i}} \Big). \tag{18}$$

Moreover, the function H is strictly convex with respect to p.

Theorem 3 *The Lagrangian is expressed as follows: for all $(x, v) \in \Omega \times \mathbb{R}^n$, one has:*

$$\begin{cases} L(x, v) = \displaystyle\sum_{i=1}^{n} \left(\sqrt{k_{off,i} \frac{v_i + d_i x_i}{d_i}} - \sqrt{k_{on,i}(x) \frac{d_i (1 - x_i) - v_i}{d_i}} \right)^2 & \text{if } v \in \Omega^v(x) \\ L(x, v) = \infty & \text{if } v \notin \Omega^v(x). \end{cases} \tag{19}$$

In addition, for all $x \in \Omega$, $L(x, v) = 0$ if and only if for all $i = 1, \ldots, n$:

$$v_i = d_i \left(\frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} - x_i \right).$$

As detailed in "Appendix C.2", we remark that the Lagrangian of the PDMP process defined in (19) is not equal to the Lagrangian of the diffusion approximation defined in "Appendix C.1", which is:

$$L_d(x, v) = \sum_{i=1}^{n} \frac{(k_{on,i}(x) + k_{off,i})^3}{4d_i^2 k_{on,i}(x) k_{off,i}} \left(v_i - d_i \left(\frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} - x_i \right) \right)^2.$$

More precisely, the Lagrangian of the diffusion approximation is a second order approximation of the Taylor expansion of the Lagrangian of the PDMP system around the velocity field associated to the deterministic limit system (4). Observe that the Lagrangian of the diffusion approximation is a quadratic mapping in v, which is expected since the diffusion approximation is described by a Gaussian process. On the contrary, the Lagrangian L given by (19) is not quadratic in v. As it had been shown in Bouchet et al. (2016) for Fast–Slow systems, this highlights the fact that the way rare events arise for the PDMP system is fundamentally different from the way they would arise if the dynamics of proteins was approximated by an SDE.

Proof of Theorem 2 Defining the 2×2 matrix

$$M^{(i)}(x, p_i) = p_i F^{(i)}(x) + K^{(i)}(x) = \begin{pmatrix} -x_i p_i d_i - k_{on,i}(x) & k_{off,i} \\ k_{on,i}(x) & -k_{off,i} + (1 - x_i) p_i d_i \end{pmatrix},$$

the Perron eigenproblem associated to $M^{(i)}$

$$M^{(i)}(x, p)\zeta^{(i)}(x, p) = H_i(x, p)\zeta^{(i)}(x, p), \qquad \zeta^{(i)} > 0,$$

implies immediately that

$$\begin{aligned} H_i(x, p) &= \frac{1}{2}\left(Tr(M^{(i)}) + \sqrt{(Tr(M^{(i)}))^2 - 4det(M^{(i)})} \right) \\ &= \frac{1}{2}\Big(p_i d_i (1 - 2x_i) - (k_{on,i}(x) + k_{off,i}) \\ &\quad + \sqrt{(p_i d_i + k_{on,i}(x) - k_{off,i})^2 + 4k_{on,i}(x) k_{off,i}} \Big). \end{aligned}$$

If we impose the constraint $\zeta_0^{(i)} + \zeta_1^{(i)} = 1$, i.e that there exists for all x, p_i, $\alpha_{p,i}(x) \in (0, 1)$ such that $\zeta^{(i)}(x, p) = \begin{pmatrix} 1 - \alpha_{p,i}(x) \\ \alpha_{p,i}(x) \end{pmatrix}$, we obtain the following equation:

$$T_x(\alpha_{p,i}(x)) = k_{on,i}(x)(1 - \alpha_{p,i}(x)) + (-k_{off,i} + (1 - x_i) p_i d_i)\alpha_{p,i}(x) = H_i(x, p)\alpha_{p,i}(x).$$

Since $T_x(0) = -k_{on,i}(x)$ and $T_x(1) = k_{off,i}$, T_x has one and only one root in $(0, 1)$. After a quick computation, one gets for all x, $p \in \Omega \times \mathbb{R}^n$:

$$\begin{cases} \alpha_{p,i}(x) = \frac{1}{2}\left(1 + \frac{\sqrt{(p_i d_i + k_{on,i}(x) - k_{off,i})^2 + 4k_{on,i}(x) k_{off,i}} - (k_{on,i}(x) + k_{off,i})}{p_i d_i} \right) & \text{if } p_i \neq 0 \\ \alpha_{p,i}(x) = \frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} & \text{if } p_i = 0. \end{cases} \tag{20}$$

Considering the tensorial structure of the problem, denoting $M_i(x, p) = p_i F_i(x) + K_i(x)$ (the tensorial version, see Sect. 2), we have by definition of M (11):

$$M = \sum_{i=1}^{n} M_i(x, p).$$

For $\zeta(x, p) = \bigotimes_{i=1}^{n} \zeta^{(i)}(x, p)$, we obtain:

$$\begin{aligned} M(x, p)\zeta(x, p) &= \sum_{i=1}^{n} \zeta^{(1)}(x, p) \otimes \cdots \otimes \underbrace{M^{(i)}(x, p)\zeta^{(i)}(x, p)}_{\text{i-th position}} \otimes \cdots \otimes \zeta^{(n)}(x, p) \\ &= \sum_{i=1}^{n} H_i(x, p_i)\zeta^{(1)}(x, p) \otimes \cdots \otimes \zeta^{(i)}(x, p) \otimes \cdots \otimes \zeta^{(n)}(x, p) \\ &= \left(\sum_{i=1}^{n} H_i(x, p_i) \right) \zeta(x, p). \end{aligned}$$

Since $\zeta > 0$, one obtains the expression (18) for the Hamiltonian:

$$H(x, p) = \sum_{i=1}^{n} H_i(x, p_i) = \sum_{i=1}^{n} p_i d_i (\alpha_{p,i}(x) - x_i). \tag{21}$$

We verify that H is strongly convex with respect to p, which follows from the following computation: for all $i, j = 1, \ldots, n$,

$$\frac{\partial^2}{\partial p_i^2} H(x, p) = \frac{2 d_i^2 k_{on,i}(x) k_{off,i}}{((p_i d_i + k_{on,i}(x) - k_{off,i})^2 + 4 k_{on,i}(x) k_{off,i})^{\frac{3}{2}}} > 0,$$

and the cross-derivatives are clearly 0. This concludes the proof of Theorem 2. □

Proof of Theorem 3 The objective is to compute the Fenchel-Legendre transform of the Hamiltonian H given by (18) in Theorem 2.

For all $x \in \Omega$ and for all $v_i \in \mathbb{R}$, the function $g : p_i \mapsto p_i v_i - H_i(x, p_i)$ is concave. An asymptotic expansion (when $p_i \to \pm\infty$) gives:

$$\begin{aligned} g(p_i) =& p_i \left(v_i - d_i \left(\frac{1}{2} (1 + \operatorname{sgn}(p_i)) - x_i \right) \right) \\ &+ \frac{1}{2} \left(k_{on,i}(x) + k_{off,i} - \operatorname{sgn}(p_i) \left(k_{on,i}(x) - k_{off,i} \right) \right) \\ &+ O\left(\frac{1}{p_i} \right). \end{aligned} \tag{22}$$

Let us study three cases. If $v_i \in (-d_i x_i, d_i(1 - x_i))$, g goes to $-\infty$ when $p_i \to \pm\infty$: thus g reaches a unique maximum in $\mathbb{R}$. At the boundary $v_i = -d_i x_i$ (resp. $v_i =$

$d_i(1 - x_i)$), g goes to $-\infty$ as p_i goes to $+\infty$ (resp. $-\infty$) and converges to $k_{on,i}(x)$ (resp. $k_{off,i}$) as p_i goes to $-\infty$ (resp. $+\infty$): then g is upper bounded and the sup is well defined. If $v_i \notin [-d_i x_i, d_i(1 - x_i)]$, $g(p_i)$ goes to $+\infty$ when either $p_i \to -\infty$ of $p_i \to +\infty$, thus g is not bounded from above.

As a consequence, $L_i(x, v_i) = \sup_{p_i} \big(p_i v_i - H_i(x, p_i)\big)$ is finite if and only if $v_i \in [-d_i x_i, d_i(1 - x_i)]$.

The Fenchel-Legendre transform of H is then given as follows: for all $x \in \Omega$ and $v \in \mathbb{R}^n$

$$L(x, v) = \sum_i L_i(x, v_i) = \sum_{i=1}^{n} \sup_{p_i \in \mathbb{R}} \big(p_i v_i - H_i(x, p_i)\big),$$

and $L(x, v)$ is finite for every $v \in \Omega^v(x)$. To find an expression for $L(x, v)$, we have to find for all $i = 1, \ldots, n$ the unique solution $p_{v,i}(x)$ of the invertible equation: $v_i = \frac{\partial H_i}{\partial p_i}(x, p_i)$. Developing the term on the right-hand side, we obtain:

$$\begin{aligned} v_i &= \frac{1}{2}\left(d_i(1 - 2x_i) + \frac{d_i(d_i p_i + k_{on,i}(x) - k_{off,i})}{\sqrt{(d_i p_i + k_{on,i}(x) - k_{off,i})^2 + 4k_{on,i}(x)k_{off,i}}}\right) \\ \iff u_i &= \frac{d_i z_i}{\sqrt{z_i^2 + c_i}}, \end{aligned} \tag{23}$$

where $u_i = 2(v_i + d_i x_i) - d_i$, $c_i = 4k_{on,i}(x)k_{off,i} > 0$, $z_i = d_i p_i + k_{on,i}(x) - k_{off,i}$.

When $v_i \in (-d_i x_i, d_i(1 - x_i))$, we have $u_i \in (-d_i, d_i)$. Thus, we obtain

$$z_i = \pm \frac{u_i \sqrt{c_i}}{\sqrt{d_i^2 - u_i^2}},$$

and as z_i and u_i must have the same sign, we can conclude for every $v_i \in (-d_i x_i, d_i(1 - x_i))$:

$$p_{v,i}(x) = \frac{1}{d_i}\left(k_{off,i} - k_{on,i}(x) + \sqrt{k_{off,i}k_{on,i}(x)}\,\frac{2(v_i + d_i x_i) - d_i}{\sqrt{(v_i + d_i x_i)(d_i(1 - x_i) - v_i)}}\right). \tag{24}$$

Injecting this formula in the expression of the Fenchel-Legendre transform, we obtain after straightforward computations:

$$\begin{aligned} L_i(x, v_i) &= p_{v,i}(x)v_i - H_i(x, p_{v,i}(x)) \\ &= \frac{1}{2}\left(\sqrt{2k_{off,i}\frac{v_i + d_i x_i}{d_i}} - \sqrt{2k_{on,i}(x)\frac{d_i(1 - x_i) - v_i}{d_i}}\right)^2. \end{aligned}$$

We finally obtain the expression when $v \in \Omega^v(x)$:

$$L(x, v) = \sum_{i=1}^{n} \left(\sqrt{k_{off,i} \frac{v_i + d_i x_i}{d_i}} - \sqrt{k_{on,i}(x) \frac{d_i(1 - x_i) - v_i}{d_i}} \right)^2 .$$

Finally, if $v \notin \Omega^v(x)$, i.e. if there exists i such that $v_i \notin [-d_i x_i, d_i(1 - x_i)]$, then $L(x, v) = L_i(x_i, v_i) = \infty$. As expected, the Lagrangian is always nonnegative. In addition, it is immediate to check that $L(x, v) = 0$ if and only if the velocity field v is the drift of the deterministic trajectories, defined by the system (4). □

5.2 Stationary Hamilton–Jacobi equation

We justified in Sect. 4.2.3 that the stationary solutions of the Hamilton–Jacobi Eq. (14) are central for finding an analytical approximation of the transition rates described in Sect. 4.1. Thus, we are going to study the existence and uniqueness (under some conditions) of functions $V \in C^1(\Omega, \mathbb{R})$ such that for all $x \in \Omega$:

$$H(x, \nabla_x V(x)) = 0. \tag{25}$$

Recalling that from (21), $H(x, p) = \sum_{i=1}^{n} p_i d_i (\alpha_{p,i}(x) - x_i)$, we construct two classes of solutions V, such that for all $i = 1, \ldots, n$, $\partial_{x_i} V(x) = 0$ or $\alpha_{\nabla_x V,i}(x) = x_i$.

The first class of solutions contains all the constant functions on Ω. From the expression (20). The second class contains all functions V such that for all $x \in \Omega$:

$$\forall i = 1, \ldots, n : \partial_{x_i} V(x) = -\frac{k_{on,i}(x)}{d_i x_i} + \frac{k_{off,i}}{d_i(1 - x_i)}. \tag{26}$$

In particular, we show in "Appendix D" that the condition (26) holds for the toggle-switch network described in "Appendix E" and studied in Sect. 6.

We will see in the next section that the class of constant solutions are associated to the deterministic system (4), which are the trajectories of convergence within the basins. We describe in Sect. 5.3.2 a more general class of solutions than (26), which defines the optimal trajectories of exit from the basins of attraction of the deterministic system.

5.3 Optimal trajectories

In the sequel we study some properties of the optimal trajectories associated to the two classes of solutions of the stationary Hamilton–Jacobi Eq. (25) introduced above.

5.3.1 Deterministic limit and relaxation trajectories

From Lemma 1, for every constant function $V(\cdot) = C$ on Ω, the associated collection of paths ϕ_t satisfying the system (16) is optimal in Ω between any pair of its points.

Replacing $p = \nabla_x V = 0$ in (16), we find that these trajectories verify at any time $t > 0$ the deterministic limit system (4):

$$\forall i \in \{1, \ldots, n\} : \dot{\phi}_i(t) = d_i \left(\frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} - \phi_i(t) \right).$$

Moreover, for every trajectory ϕ_t solution of this system, we have for any $T > 0$:

$$J_T(\phi_t) = \int_0^T L(\phi_t, \dot{\phi}_t) dt = V(\phi_T) - V(\phi_0) = 0. \tag{27}$$

We call such trajectories the *relaxation trajectories*, as they characterize the optimal path of convergence within every basin. From Theorem 3, these relaxation trajectories are the only zero-cost admissible trajectories.

5.3.2 Quasipotential and fluctuation trajectories

We now characterize the optimal trajectories of exit from the basins. We are going to show that the condition (C) defined below is sufficient for a solution $V \in C^1(\Omega, \mathbb{R})$ of the Eq. (25) to define optimal trajectories realizing the rare events described by the probabilities (7).

Definition 3 We define the following condition on a function $V \in C^1(\Omega, \mathbb{R})$:

(C) The set $\{x \in \Omega \mid \nabla_x V(x) = 0\}$ is reduced to isolated points.

The results presented below in Theorem 4 are mainly an adaptation of Theorem 3.1, Chapter 4, in Freidlin and Wentzell (2012). In this first Theorem, we state some properties of solutions $V \in C^1(\Omega, \mathbb{R})$ of (25) satisfying the condition (C):

Theorem 4 *Let $V \in C^1(\Omega, \mathbb{R})$ be a solution of (25).*

(i) *For any optimal trajectory ϕ_t satisfying the system (16) associated to V, for any time t we have the equivalence:*

$$\left(\forall i \in \{1, \ldots, n\}, \ \phi_i(t) = \frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} \right) \iff \dot{\phi}(t) = 0.$$

(ii) *The condition (C) implies that the gradient of V vanishes only on the stationary points of the system (4).*

(iii) *If V satisfies (C), then V is strictly increasing on any trajectory which solves the system (16), such that the initial condition is not an equilibrium point of the system (4). Moreover, for any basin of attraction Z_i associated to an attractor X_{eq,Z_i}, we have:*

$$\forall x \in Z_i \setminus X_{eq,Z_i}, \ V(x) > V(X_{eq,Z_i}).$$

(iv) *If V satisfies the condition (C), under the assumption that* $\lim_{x\to\partial\Omega} V(x) = +\infty$, *we have the formula:*

$$\forall x \in \Omega,\ V(x) = \min_{\{a\in\Omega | \nabla_x V(a)=0\}} V(a) + Q(a,x).$$

(v) *Let us consider $V, \tilde{V} \in C^1(\Omega, \mathbb{R})$ two solutions of (25) satisfying the condition (C). The stable equilibria of the system defined for every time t by $\dot{\phi}_t = -\nabla_p H(\phi_t, \nabla_x V(\phi_t))$ are exactly the attractor of the deterministic system (4) $(X_{eq,Z_i})_{Z_i\in Z}$. We denote $(Z_i^f)_{Z_i\in Z}$ the basins of attraction which are associated to these equilibria: at least on $\bigcup_{Z_i\in Z} \bar{Z}_i^f$, the relation $\nabla_x V = \nabla_x \tilde{V}$ is satisfied.*
Moreover, under the assumptions 1. that $\lim_{x\to\partial\Omega} V(x) = \lim_{x\to\partial\Omega} \tilde{V}(x) = \infty$, and 2. that between any pair of basins (Z_i^f, Z_j^f), we can build a serie of basins $(Z_{u_k}^f)_{k=1,\dots,m}$ such that $u_0 = 1$, $u_m = j$ and for all $k < m$, $\bar{Z}_{u_k}^f \cap \bar{Z}_{u_{k+1}}^f \neq \varnothing$, then V and $\tilde{V}$ are equal in Ω up to a constant.

Note that the point (iii) makes these solutions consistent with the interpretation of the function V as the rate function associated to the stationary distribution of the PDMP system, presented in Sect. 4.2.2. Indeed, as every random path converges in probability when $\varepsilon \to 0$ to the solutions of the deterministic system (4) (Faggionato et al. 2009), the rate function has to be minimal on the attractors of this system, which then corresponds to the points of maximum likelihood at the steady state. It should also converge to $+\infty$ on $\partial\Omega$, as the cost of reaching any point of the boundary is infinite (see Corollary 2, in the proof of Theorem 5). However, we see in (v) that the uniqueness, up to a constant, needs an additional condition on the connection between basins which remains not clear for us at this stage, and which will be the subject of future works.

If $V \in C^1(\Omega, \mathbb{R})$ is a solution of (25) saitsfying (C), we call a trajectory solution of the system (16) associated to V a *fluctuation trajectory*.

We observe that any function satisfying the relation (26) belongs to this class of solutions of (25), and then that in particular, such C^1 function exists for the toggle-switch network. In that special case, we can explicitly describe all the fluctuation trajectories: for any time t, replacing $p_i(t) = -\frac{k_{on,i}(\phi(t))}{d_i\phi_i(t)} + \frac{k_{off,i}}{d_i(1-\phi_i(t))}$ in the system (16), we obtain

$$\forall i \in \{1,\dots,n\} :\ \dot{\phi}_i(t) = d_i \left(\frac{k_{off,i}\phi_i(t)^2}{k_{on,i}(\phi(t))(1-\phi_i(t))^2 + k_{off,i}\phi_i(t)^2} - \phi_i(t) \right). \tag{28}$$

In the second theorem (Theorem 5), we justify that the fluctuation trajectories are the optimal trajectories of exit:

Theorem 5 *Let us assume that there exists a function $V \in C^1(\Omega, \mathbb{R})$ which is solution of (25) and satisfies the condition (C). For any basin $Z_i \in Z$, there exists at least one basin Z_j, $j \neq i$, such that there exists a couple (x_0, ϕ_t), where $x_0 \in \Gamma_{Z_i}$ and ϕ_t is a fluctuation trajectory, and such that $\phi(0) = x_0$, $\phi(T) \underset{T\to\infty}{\to} x^{ij} = \underset{y\in\partial Z_i\cap\partial Z_j}{argmin}\ V(y)$.*

Let us denote $X_{un}^{ij} = \{x \in \partial Z_i \cap \partial Z_j \mid \nabla_x V(x) = 0\}$, $x_{un}^{ij} = \underset{y\in X_{un}^{ij}}{argmin}\, V(y)$ and $R_{ij} = \bigcup_{k\neq i,j} \{\partial Z_i \cap \partial Z_k\}$. Under the following assumption

(A) *any relaxation trajectory starting in $\partial Z_i \cap \partial Z_j$ stays in $\partial Z_i \cap \partial Z_j$,*

we have $x^{ij} = x_{un}^{ij}$ and:

$$Q_{R_{ij}}(X_{eq,Z_i}, \partial Z_i \cap \partial Z_j) = V(x_{un}^{ij}) - V(X_{eq,Z_i}).$$

In particular, if there exists a fluctuation trajectory between any attractor X_{eq,Z_i} and every saddle points of the deterministic system (16) on the boundary ∂Z_i, and if the assumption (A) of Theorem 5 is verified for every basin Z_j, $j \neq i$, the function V allows to quantify all the optimal costs of transition between the basins. This is generally expected because the attractors are the only stable equilibria for the reverse fluctuations (see the proof of Theorem 4.(v)). The proofs of Theorems 4 and 5 use classical tools from Hamiltonian system theory and are postponed to Appendix H.

When a solution $V \in C^1(\Omega, \mathbb{R})$ satisfying (C) exists, the saddle points of the deterministic system (4) are then generally the bottlenecks of transitions between basins and the function V characterizes the energetic barrier between them. The function $Q(X_{eq,Z_i}, \cdot)$ depends on the basin Z_i, which is a local property: it explains why the function V is generally called the global quasipotential, and $Q(X_{eq,Z_i}, \cdot)$ the local quasipotential of the process (Zhou and Li 2016).

The precise analysis of the existence of a regular solution V satisfying (C) for a given network is beyond the scope of this article. When it is impossible to find a regular solution, more general arguments developed within the context of Weak KAM Theory can allow to link the viscosity solutions of the Hamilton–Jacobi equation to the optimal trajectories in the gene expression space (Fathi 2008).

5.4 Partial conclusion

We have obtained in Theorem 3 the form of the Lagrangian in the variational expression (12) for the rate function J_T associated to the LDP for the PDMP system (1). We have also highlighted the existence and interpretation of two types of optimal trajectories.

The first class consists in relaxation trajectories, which characterize the convergence within the basins. The fact that they are the only trajectories which have zero cost justifies that any random path X_t^ε converges in probability to a relaxation trajectory.

When there exists a function $V \in C^1(\Omega, \mathbb{R})$ satisfying (26), the system (28) defines the second class of optimal trajectories, called the fluctuation trajectories. From Theorem 5, for every basin Z_i, there exists at least one basin Z_j, $j \neq i$, and a trajectory Φ_t^{ij} which verifies this system, starts on $x_0 \in \Gamma_{min,Z_i}$ and reaches a point of $x \in \partial Z_i \cap \partial Z_j$

such that $Q(x_0, x) = Q(x_0, \partial Z_i \cap \partial Z_j)$. This trajectory then realizes the rare event of probability $p_{ij}(x_0)$. Injecting the velocity field defining (28) in the Lagrangian (19), we deduce:

$$-\varepsilon \ln\left(p_{ij}^{\varepsilon}(x_0)\right) \underset{\varepsilon\to 0}{\to} J_T(\Phi_t^{ij}) = \int_0^T \sum_{i=1}^{n} \frac{\left(k_{on,i}(\phi(t))(1-\phi_i(t)) - k_{off,i}\phi_i(t)\right)^2}{k_{on,i}(\phi(t))(1-\phi_i(t))^2 + k_{off,i}\phi_i(t)^2} dt. \tag{29}$$

If the assumption (A) of Theorem 5 is verified, this minimum is necessarily reached on a saddle point of V on $\partial Z_i \cap \partial Z_j$ and in that case, the time T must be taken infinite. Then, the formula (29) can be injected in the approximation (17), and the method described in Sect. 4.2.2 allows to compute the probability of the form (7) for the pair (Z_i, Z_j).

Moreover, for every basin Z_k, $k \neq i$, if the assumption (A) of Theorem 5 is verified and if there exists, for any saddle point $x_{un}^{ik} \in \partial Z_i \cap \partial Z_k$, a trajectory satisfying the system (28) which starts at x_0 and reaches x_{un}^{ik} (at $T \to \infty$), the formula (29) can also be injected in the approximation (17) for the pair (Z_i, Z_k), and the method described in Sect. 4.2.3 allows then to compute the probabilities of the form (7) for any pair of basins $(Z_i, Z_k)_{k=1,\ldots,m}$.

6 Application to the toggle-switch network

In this section, we consider the class of interaction functions defined in Appendix D for a network with two genes ($n = 2$). This function comes from a chromatin model developed in Herbach et al. (2017) and is consistent with the classical Hill function characterizing promoters switches. Using results and methods described in the previous sections, we are going to reduce the PDMP system when the GRN is the toggle-switch network described in "Appendix E". After defining the attractors of the deterministic system (4), building the optimal fluctuation trajectories between these attractors and the common boundaries of the basins, we will compute the cost of the trajectories and deduce, from the approximation (17), the transition probabilities of the form (7) as a function of ε, up to the prefactor. We will compute these probabilities for some ε with the AMS algorithm described in "Appendix G" for obtaining the prefactor. We will then approximate the transition rates characterizing the discrete Markov chain on the cellular types, given by the formula (8), for many values of ε. We will finally compare these results to the ones given by a crude Monte-Carlo method.

6.1 Computation of the attractors, saddle points and optimal trajectories

First, we compute the stable equilibrium points of the PDMP system (1). The system (5) has no explicit solution. We present a simple method to find them, which consists in sampling a collection of random paths in Ω: the distribution of their final position after a long time approximates the marginal on proteins of the stationary distribution. We use these final positions as starting points for simulating the relaxation trajectories,

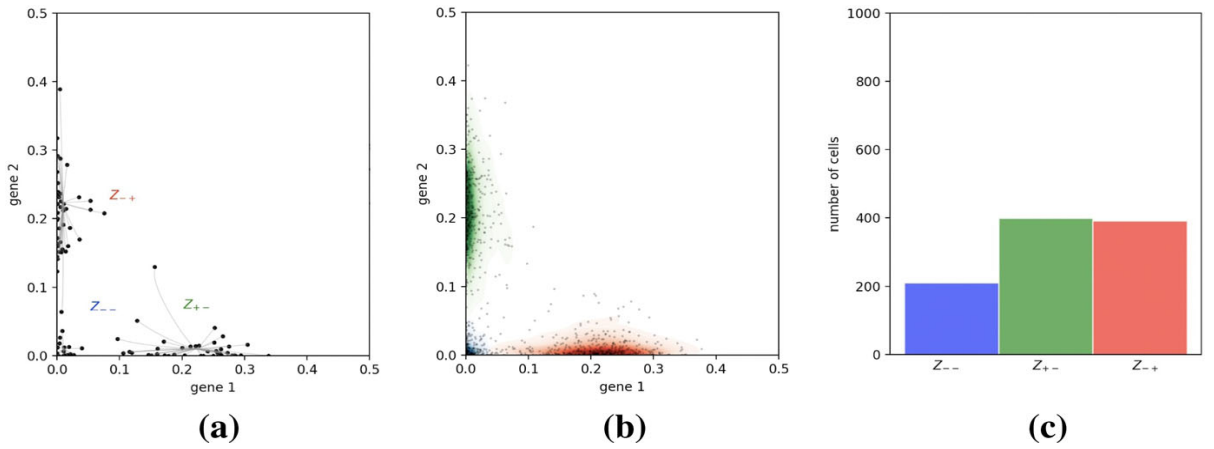


Fig. 5 **a** 100 cells are plotted under the stationary distribution. The relaxation trajectories allow to link every cell to its associated attractor. **b** 1000 cells are plotted under the stationary distribution. They are then classified depending on their attractor, and this figure sketches the kernel density estimation of proteins within each basin. **c** The ratio of cells that are found within each basin gives an estimation of the stationary distribution on the basins

described by (4), with an ODE solver: each of these relaxation trajectories converges to one of the stable equilibrium points. This method allows to obtain all the stable equilibrium corresponding to sufficiently deep potential wells (see Fig. 5). Possible other potential wells can be omitted because they correspond to basins where the process has very low probability of going, and which do not impact significantly the coarse-grained Markov model.

Second, we need to characterize the fluctuation trajectories. In "Appendix D", we introduced the interaction function and proved that for any symmetric two-dimensional network defined by this function, i.e such that for any pair of genes (i, j), $\theta_{ij} = \theta_{ji}$ (where θ is the matrix characterizing the interactions between genes), there exists a function V such that the relation (26) is verified. This is then the case for the toggle-switch network, which is symmetric. We have proved in Sect. 5.2 that such function V solves the Hamilton–Jacobi Eq. (25), and verifies the condition (C). Thus, the system (28) defines the fluctuation trajectories.

Third, we need to find the saddle points of the system (4). As we know that for any attractor, there exists at least one fluctuation trajectory which starts on the attractor and reaches a saddle point (in an infinite time), a naive approach would consist in simulating many trajectories with different initial positions around every attractors, until reaching many saddle points of the system. This method is called a shooting method and may be very efficient in certain cases. But for the toggle-switch, we observe that the fluctuation trajectories are very unstable: this method does not allow to obtain the saddle points.

We develop a simple algorithm which uses the nonnegative function $l(\cdot) = L(\cdot, \nu_v(\cdot))$, which corresponds to the Lagrangian evaluated on the drift ν_v of the fluctuation trajectories defined by the system (28). We have :

$$l : x \to L(x, \nu_v(x)) = \sum_{i=1}^{n} \frac{\left(k_{on,i}(x)(1-x) - k_{off,i}x\right)^2}{k_{on,i}(x)(1-x)^2 + k_{off,i}x^2}.$$

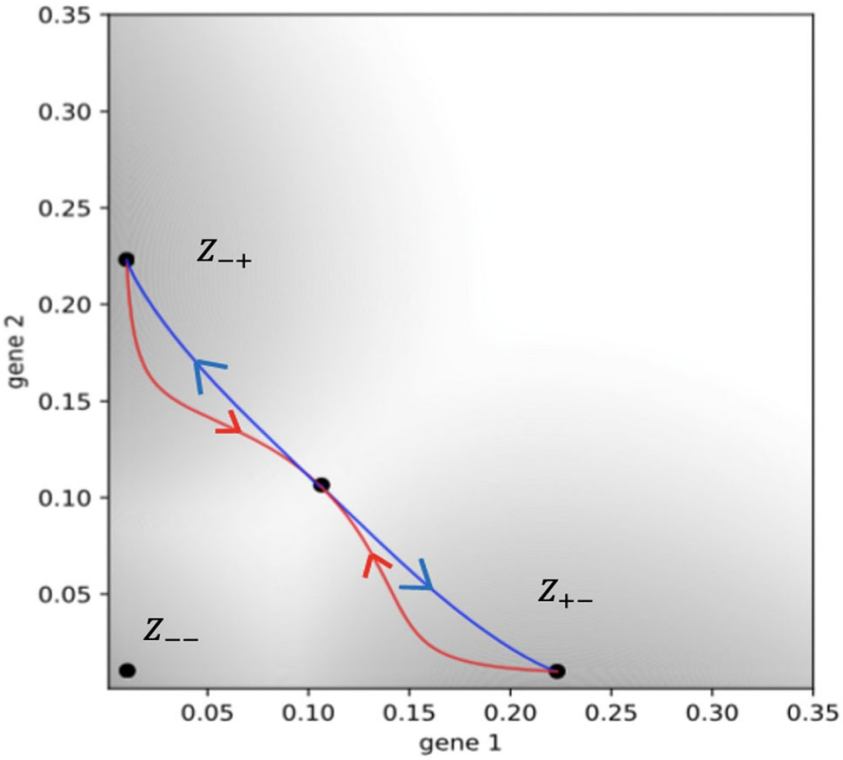


Fig. 6 The optimal fluctuation trajectories from a first attractor continued by the relaxation trajectories reaching the second attractor, for the pair (Z_{+-}, Z_{-+}). We omit the other pairs of attractors (Z_{+-}, Z_{--}) and (Z_{--}, Z_{-+}), because their optimal trajectories are simply straight lines (colour figure online)

As expected, since ν_v cannot be equal to the drift of a relaxation trajectory except on the stationary points of the relaxation trajectories and since the Lagrangian $L(x, v)$ is equal to 0 if and only if v corresponds to the drift of a relaxation trajectory, the function l vanishes only on these stationary points. If there exists a saddle point connecting two attractors, this function will then vanish there. The algorithm is described in Appendix I. For the toggle-switch network, it allows to recover all the saddle points of the system (4).

Fourth, we want to compute the optimal trajectories between every attractors and the saddle points on the boundary of its associated basin. Using the reverse of the fluctuation trajectories, for which the attractors of the system (4) are asymptotically stable (see the proof of Theorem 4.(v)), we can successively apply a shooting method around every saddle points. We observe that for the toggle-switch network, for any saddle point at the boundary of two basins, there exists a reverse fluctuation trajectory which converges to the attractors of both basins. For any pair of basins (Z_i, Z_j), we then obtain the optimal trajectories connecting the attractor X_{eq,Z_i} and the saddle points belonging to the common boundary $\partial Z_i \cap \partial Z_j$ (see Fig. 6).

Finally, we want to compute the optimal transition cost between any pair of basins (Z_i, Z_j). We observe that every relaxation trajectories starting on the common boundary of two basins stay on this boundary and converge to a unique saddle point inside: the assumption (A) of Theorem 5 is then verified. It follows from this theorem that the optimal trajectory between any basin Z_i and Z_j necessarily reaches $\partial Z_i \cap \partial Z_j$ on a saddle point, and then that the optimal transition cost is given by the trajectory which

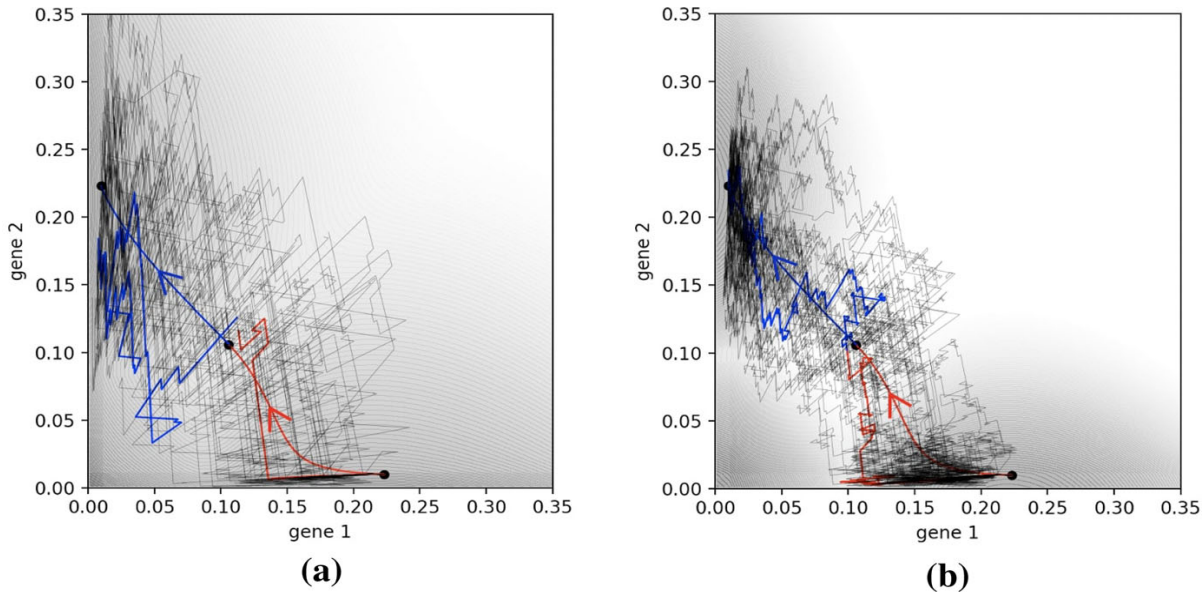


Fig. 7 Comparison between the optimal trajectory of the Fig. 6 and 30 random paths conditioned on reaching, from a point of the boundary of the R-neighborhood of an attractor $X_{eq,Z_{-+}}$, the r-neighborhood of a new attractor $X_{eq,Z_{+-}}$ before the r neighborhood of the first attractor $X_{eq,Z_{-+}}$, with $r < R$. We represent this comparison for **a** $\varepsilon = 1/7$ and **b** $\varepsilon = 1/21$. For each figure, one of these random paths is colored, separating the fluctuation and the relaxation parts (colour figure online)

minimizes the cost among all those found previously between the attractor and the saddle points. We denote this optimal trajectory ϕ_t^{ij}. Its cost is explicitly described by the formula (29) (with $T \to \infty$), which is then the optimal cost of transition between Z_i and Z_j.

The LDP ensures that for all $\delta, \eta > 0$, there exists ε' such that for all $\varepsilon \in (0, \varepsilon')$, a random path X_t^ε reaching Z_j from Γ_{Z_i} before γ_{Z_i}, verifies: $\sup_t\{|\ X_t^\varepsilon - \phi_t^{ij}\ |\} \leq \delta$ with probability larger than $1 - \eta$. In other words, given a level of resolution δ, we could then theoretically find ε such that any trajectory of exit from Z_i to Z_j would be indistinguishable from trajectory ϕ_t^{ij} at this level of resolution. But in practice, the event $\{\tau_{Z_j}^\varepsilon < \tau_{\gamma_{Z_i}}^\varepsilon\}$ is too rare to be simulated directly for such ε.

We plot in Fig. 7 two sets of random exit paths, simulated for two different ε, illustrating the fact that the probability of an exit path to be far from the optimal fluctuation trajectory decreases with ε.

6.2 Comparison between predictions and simulations

For each pair of basins (Z_i, Z_j), the expression (29) provides an approximation of the probability of the rare event $\{\tau_{Z_j}^\varepsilon < \tau_{\gamma_{Z_i}}^\varepsilon\}$, up to a prefactor, and the approximation (17) allows to deduce the associated transition rate. We plot in Fig. 8 the evolution of these two estimations, as ε decreases, comparing respectively to the probabilities given by the AMS algorithm and the transition rates computed with a Monte-Carlo method. As in Bréhier and Lelièvre (2019), we decide to plot these quantities in logarithmic scale. We observe that, knowing the prefactor, the Large deviations approximation is accurate even for $\varepsilon > 0.1$, and that induced transition rates are close to the ones

observed with a Monte-Carlo method too. We represent in Fig. 10b the variance of the estimator of the transition rates given by the AMS method.

We also remark that our analysis provides two ways of estimating the stationary measure of the discrete coarse-grained model. On the one hand, we can obtain a long-time proteins distribution of thousands of cells by simulating the PDMP system (1) from random initial conditions: by identifying each cell with a basin, as shown in Fig. 9a, we can find a vector μ_b describing the ratio of cells belonging to each basin. When the number and length of the simulations are large enough, this vector μ_b should be a good approximation of the stationary measure on the basins. On the other hand, the transition rates allows to build the transition matrix M of the discrete Markov process on the basins, $\hat{Z}_l^\varepsilon$, defined in Sect. 3.2. If the exponential approximation of the first passage time from every basin is accurate, then the stationary distribution on the basins should be well approximate by the unique probability vector such that $\mu_z M = 0$ (see Fig. 9b).

Monte-Carlo methods for approximating the transition rates have a very high computational cost when ε is small. Thus, comparing these two stationary distributions appears as a good alternative for verifying the accuracy of the transition rates approximations. We plot in Fig. 10a the evolution of the total variation distance between these two stationary distributions as ε decreases. We observe that the total variation is small even for realistic values of ε. The variance of the estimator μ_b is very small (given it is estimated after a time long enough) but the estimator μ_z accumulates all numerical errors coming from the estimators needed to compute the transition rates: this is likely to explain the unexpected small increases observed in this curve for $\varepsilon = 1/6$. We represent in Fig. 10b the variance of the transition rates estimators between every pair of attractors used for estimating the distribution μ_z in Fig. 10a, for $\varepsilon = 1/7$: as expected, this variance increases with the transition rates.

The similarity between the two distributions μ_z and μ_b seems to justify the Markovian approximation of the reduced process $\hat{Z}_t^\varepsilon$ for small but realistic ε: at least for the toggle-switch network, the coarse-grained model, evolving on the basins of attractions seen as cellular types, describes accurately the complex behaviour of a cell in the gene expression space.

6.3 Applicability for more complex networks

It is in general very complex to find a solution $V \in C1(\Omega, \mathbb{R})$ to the stationary Hamilton–Jacobi Eq. (25) which satisfies the condition (C) for general networks, when the number of genes is greater than 2. In order to apply the strategy developed in Sect. 5, for computing the cost of the optimal trajectories of transition between two basins, it would be then necessary to build a computational method for approximating such solution. Although the most common approach in this case consists in finding optimal trajectories without computing a potential (see Heymann and Vanden-Eijnden 2008 or Li et al. 2021 for more recent works), some methods have been recently built for SDEs model, like Langevin dynamics (Brackston et al. 2018). Such computational method for the PDMP system is beyond the scope of the article. However, we remark that even if there are no reasons for the trajectories satisfying the system (28) to be

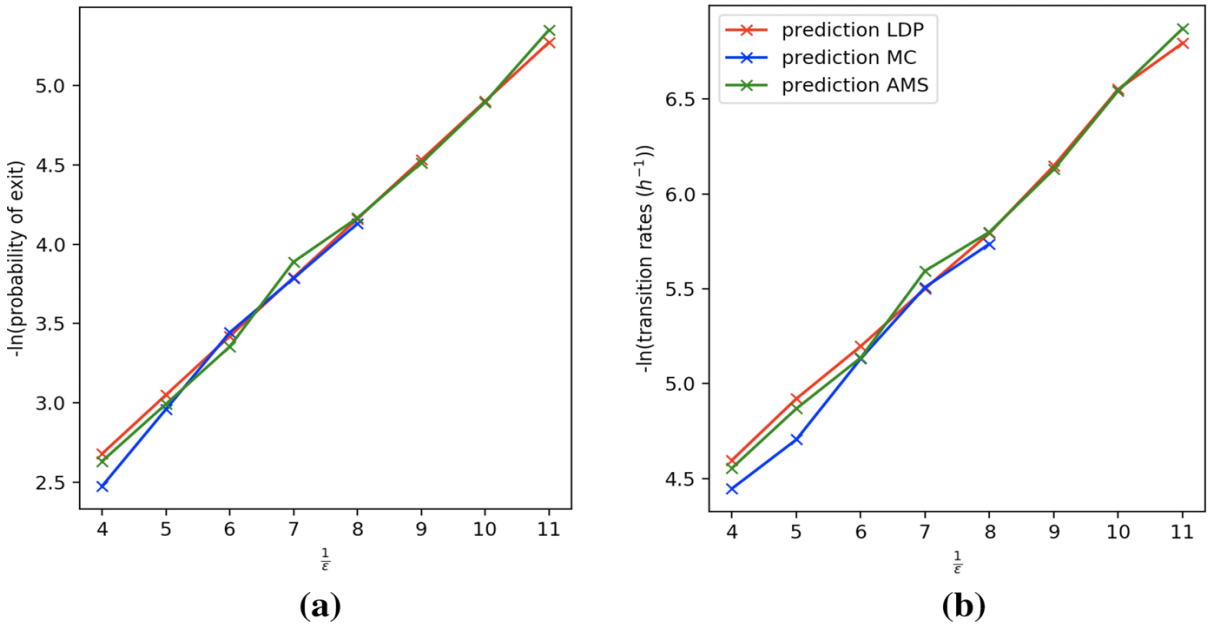


Fig. 8 **a** Comparison between the probabilities (7) between the basins Z_{+-} and Z_{--}, in logarithmic scale, given by the Large deviations approximation (in red) and the AMS algorithm (in green). The prefactor is computed for $\varepsilon = 1/8$ and the red curve is then adjusted to fit the numerical results. The blue curve corresponds to the probabilities obtained with a Monte-Carlo method. **b** Comparison between the transition rates between the basins Z_{+-} and Z_{--}, in logarithmic scale, given by the formula (8), where the probability (7) is given by the Large deviations approximation (in red) and the AMS algorithm (in green). The blue curve corresponds to the transition rates obtained with a Monte-Carlo method, by the formula (6). The quantities obtained by a Monte-Carlo method, in blue, are not represented after $\varepsilon = 1/8$ because the transition rates become too small to be efficiently computed (colour figure online)

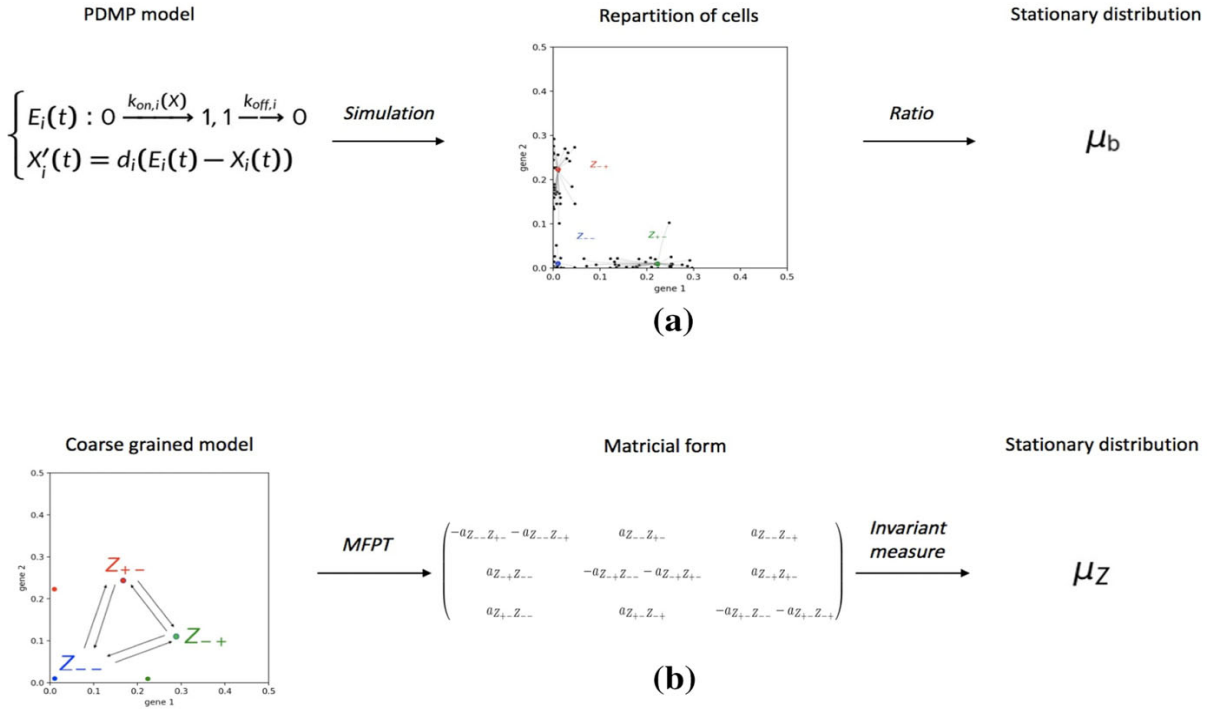


Fig. 9 Comparison between the two methods for obtaining estimators of the stationary distributions on the basins: μ_b (**a**) and μ_z (**b**)

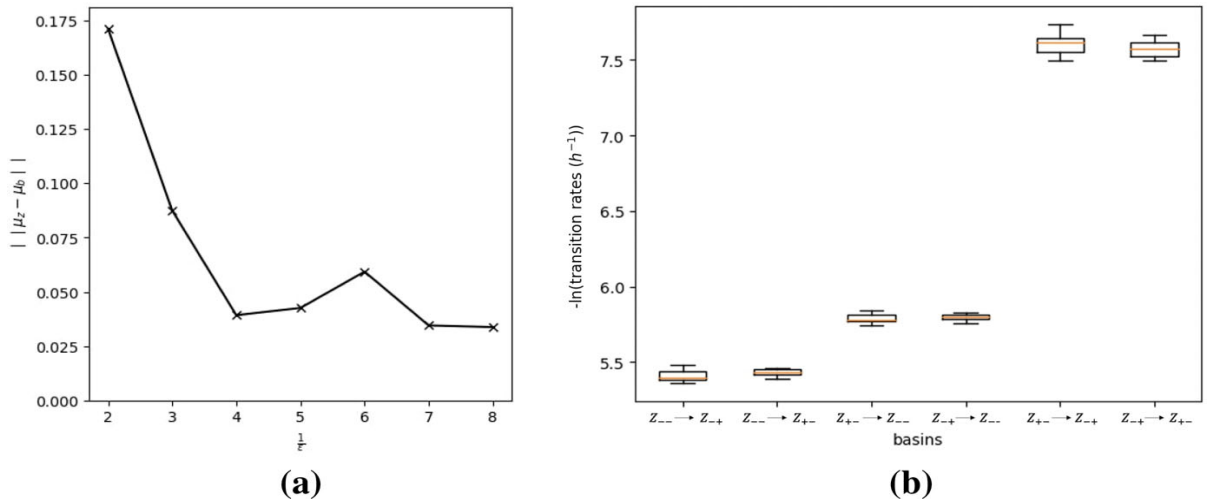


Fig. 10 **a** The total variation of the difference between μ_b and μ_z as a function of ε^{-1}. **b** Boxplots representing the variation of the transition rates for 10 iterations of the method used in **a**, between each pair of basins for $\varepsilon = 1/7$

optimal when no function satisfying the relation (26) can be found, our computational method still allows to compute these trajectories, and we observe that they generally still bound the attractors and the saddle points of the deterministic system (4). Their costs can then be used as a proxy for the probabilities of the form (7): we observe in Figs. 16b and 17b in Appendix J that for two non-symmetric networks of respectively 3 and 4 genes, our method still provides good results.

7 Discussion

Using the WKB approximation presented in Sect. 4.2.2 and the explicit formulas for the Hamiltonian and the Lagrangian detailed in Sect. 5.1, we are going now to analyze more precisely how the LDP for the proteins can be interpreted in regards to the dynamics of promoters, and we will see how two classical notions of energies can be interpreted in light of this analysis.

7.1 Correspondences between velocities and promoters frequency lead to energetic interpretations

The main idea behind the LDP principle for the PDMP system is that a slow dynamics on proteins coupled to the fast Markov chain on promoters rapidly samples the different states of P_E according to some probability measure $\pi = (\pi_e)_{e \in P_E}$. The value $\sum_{e, e_i = 1}^{n} \pi_e$ corresponds then to the parameter of the Bernoulli describing the random variable E_i, and can be interpreted as the frequency of the promoter of gene i.

The point of view of Faggionato et al. (2009) consisted in stating a LDP for the PDMP system by studying the deviations of π from the quasistationary distribution (3). The work of Bressloff and Faugeras (2017) consists in averaging the flux asso-

ciated to the transport of each protein over the measure π, in order to build a new expression of this LDP which depends only on the protein dynamics. Its coupling with an Hamiltonian function through a Fenchel-Legendre transform allows to apply a wide variety of analytical tools to gain insight on the most probable behaviour of the process, conditioned on rare events. In this Section, we see how correspondences between these different points of view on the LDP shed light on the meaning of the Hamiltonian and Lagrangian functions and lead to some energetic interpretations.

7.1.1 Correspondence between velocity and promoter frequency

Let us fix the time. The velocity field of the PDMP system, that we denote Φ, is a n-dimensional vector field, function of the random vectors E, X, which can be written for any $i = 1, \dots, n$:

$$\Phi_i = (1 - E_i) \times v_{off,i}(X_i) + E_i \times v_{on,i}(X_i), \tag{30}$$

with the functions $v_{off,i} : x_i \mapsto -d_i x_i$ and $v_{on,i} : x_i \mapsto d_i(1 - x_i)$ for any $x \in \Omega$.

For all $i = 1, \dots, n$, let $\rho_i : x \mapsto \mathbb{E}(E_i \mid X = x)$ denote the conditional expectation of the promoter E_i knowing a protein vector X. As presented in Sect. 3.1, the quasistationary approximation identifies the vector field ρ to the invariant measure of the Markov chain on the promoter states.

For a given conditional expectation of promoters ρ, the vector field $v_\rho : x \mapsto \mathbb{E}_{E\sim\rho(x)}(\Phi \mid X = x)$ is defined for all $x \in \Omega$ by:

$$\begin{aligned} \forall i = 1, \dots, n,\ v_{\rho,i}(x) &= (1 - \rho_i(x))v_{off,i}(x) + \rho_i(x)v_{on,i}(x) \\ &= d_i(\rho_i(x) - x_i) \in [-d_i x_i, d_i(1 - x_i)]. \end{aligned} \tag{31}$$

Denoting Ω^v the set of vector fields v continuous on Ω, such that for all $x \in \Omega$, $v(x) \in \Omega^v(x)$, we see that $v_\rho \in \Omega^v$. Conversely, the formula (31) can be inverted for associating to every velocity field $v \in \Omega^v$, characterizing the protein dynamics, a unique conditional expectation of promoters states knowing proteins, ρ_v, which is the unique solution to the reverse problem $v(\cdot) = \mathbb{E}_{E\sim\rho(\cdot)}(\Phi \mid X = \cdot)$, and which is defined by:

$$\forall x \in \Omega,\ \forall i = 1, \dots, n :\ \rho_{v,i}(x) = \frac{v_i(x) - v_{off,i}(x)}{d_i} \in [0, 1]. \tag{32}$$

7.1.2 Dynamics associated to a protein field

We detailed above the correspondence between any admissible velocity field $v \in \Omega^v$ and a unique vector field ρ_v describing a conditional expectation of promoters states knowing proteins. Moreover, the proof of Theorem 2 reveals that for any vector field $p : \Omega \mapsto \mathbb{R}^n$, we can define a unique vector field $\alpha_p : \Omega \mapsto (0, 1)^n$ by the expression (20).

As presented in Sect. 4.2.2, we denote V the leading order term of the Taylor expansion in ε of the function S defined in (13), such that the distribution of the

PDMP system is defined at a fixed time t, and for all $e \in P_E$, by $u_e(\cdot) = \pi_e(\cdot)e^{-\frac{S(\cdot)}{\varepsilon}}$, where $\pi(x)$ is a probability vector in S_E for all $x \in \Omega$.

On the one hand, we have seen in Sect. 4.2.2 that for all $x \in \Omega$, the eigenvector $\zeta(x, \nabla_x V(x))$ of the spectral problem (10) (for $p = \nabla_x V(x)$) corresponds to the leading order term of the Taylor expansion in ε of $\pi(x)$. For all $i = 1, \ldots, n$, the quantity $\sum_{e \in P_E, e_i = 1} \zeta_e(x, \nabla_x V(x))$ then represents the leading order approximation of the conditional expectation $\rho_i(x) = \mathbb{E}(E_i \mid X = x)$. On the other hand, if we denote the gradient field $p = \nabla_x V$ defined on Ω, we recall that for all $x \in \Omega$: $\zeta(x, p) = \bigotimes_{i=1}^{n} \begin{pmatrix} 1 - \alpha_{p,i}(x) \\ \alpha_{p,i}(x) \end{pmatrix}$. We then obtain:

$$\alpha_{p,i}(x) = \sum_{e \in P_E, e_i = 1} \zeta_e(x, \nabla_x V(x)) \approx \rho_i(x).$$

This interpretation of the vector α_p, combined with the relation (32), allows us to state that the velocity field defined for all $x \in \Omega$ by $v_{\alpha_p}(x) = \big(d_i(\alpha_{p,i}(x) - x_i)\big)_{i=1,\ldots,n} \in \Omega^v(x)$ characterizes, in the weak noise limit, the protein dynamics associated to the proteins distribution $u = e^{-\frac{S(\cdot)}{\varepsilon}}$.

We see that the velocity field v_{α_p} corresponds to the drift of the deterministic system (4) if and only if $\alpha_p = \frac{k_{on}}{k_{on}+k_{off}}$, and then if and only if $p = 0$ (see Sect. 5.2). The gradient field p can be understood as a deformation of the deterministic drift, in the weak noise limit.

We recall that for all $p \in \mathbb{R}^n$, we have from (21):

$$H(x, p) = \sum_{i=1}^{n} p_i d_i(\alpha_{p,i}(x) - x_i).$$

With the previous notations, the Lagrangian associated to a velocity field v can then be written on every $x \in \Omega$ as a function of α_p and ρ_v:

$$\begin{aligned} L(x, v(x)) &= \sum_{i=1}^{n} p_i(x) d_i(\rho_{v,i}(x) - x_i) - \sum_{i=1}^{n} p_i(x) d_i(\alpha_{p,i}(x) - x_i) \\ &= \langle p(x), v(x) - v_{\alpha,p}(x) \rangle, \end{aligned}$$

where $p(x) = p_v(x)$ is defined by the expression (24). Thus, we see that the duality between the Lagrangian and the Hamiltonian, that we intensively used in this article for analyzing the optimal trajectories of the PDMP system, and which is expressed through the relation (24) between the variables v and p, also corresponds to a duality between two promoters frequencies ρ_v and α_p associated to the velocity fields v and v_{α_p}.

The situation is then the following: for a given proteins distribution $u(\cdot) = e^{-\frac{S(\cdot)}{\varepsilon}}$ such that the first order approximation of S in ε, V, is differentiable on Ω, the velocity

field v associated by duality to the gradient field $p = \nabla_x V$, and which characterizes a collection of optimal trajectories of the PDMP system (satisfying the system (16) associated to V) when u is the stationary distribution, does not correspond to the protein velocity v_{α_p} associated to the distribution u in the weak noise limit, except when the Lagrangian vanishes on (x, v). Alternatively speaking, the optimal trajectories associated to a distribution in the sense of Large deviations, characterized by the velocity field v, do not correspond to the trajectories expected in the weak noise limit, characterized by the velocity field v_{α_p}. This is an important limit for developing a physical interpretation of the Hamiltonian system in analogy with Newtonian mechanics. However, the correspondence between promoters states distributions and velocity fields developed above leads us to draw a parallel with some notions of energy.

7.1.3 Energetic interpretation

Following a classical interpretation in Hamiltonian system theory, we introduce a notion of energy associated to a velocity field:

Definition 4 Let us consider $x \in \Omega$ and $v \in \Omega^v$. The quantity $E_v(x) = H(x, p_v(x))$ is called the energy of the velocity field v on x, where $p_v(x)$ is defined by the expression (24).

Interestingly, combining the expression of the Hamiltonian given in Theorem 2 with the expressions (24) and (32), the energy of a velocity v on every $x \in \Omega$ can be rewritten:

$$E_v(x) = \sum_{i=1}^{n} \frac{\sqrt{k_{on,i}(x)k_{off,i}}}{d_i} \left(\frac{\mathbb{E}_{\rho_{v,i}}(|\Phi_i| \mid X = x)}{\sigma(\rho_{v,i}(x))} - \frac{\mathbb{E}_{\overline{\rho}_i}(|\Phi_i| \mid X = x)}{\sigma(\overline{\rho}_i(x))} \right)$$

where for all $i = 1, \ldots, n$, Φ_i is the random variable defined by the expression (30), which follows, conditionally to proteins, a Bernoulli distribution of parameter $\rho_{v,i}$, and $\sigma(\rho_{v,i}(x)) = \sqrt{\rho_{v,i}(x)(1 - \rho_{v,i}(x))}$ denotes its standard deviation.

Finally, we have $\mathbb{E}_{\rho_{v,i}}(|\Phi_i| \mid X = x) = (1 - \rho_{v,i}(x))v_{off,i}(x) + \rho_{v,i}(x)v_{on,i}(x)$, and $\overline{\rho}$ denotes the quasistationary distribution described in (3).

Formally, the energy of a promoter distribution can then be decomposed in two terms : a first term describing its velocity in absolute terms, scaled by its standard deviation, and a second term depending on the network. A high energy distribution on a point x is characterized by a fast and deterministic protein dynamics in regards respectively to the velocity of the quasistationary approximation on x and the standard deviation of its associated promoter distribution.

We remark that this notion of energy does not depend on the proteins distribution, but only on the promoters frequency ρ_v around a certain location x. Depending on x only through the vector field ρ_v (and the functions $k_{on,i}$), it is likely to be interpreted as the kinetic energy of a cell.

The potential $V = -\ln(\hat{u})$, where $\hat{u}$ is the marginal on proteins of the stationary distribution of the stochastic process, is classically interpreted as a notion of potential energy, not depending on the effective promoter frequency. Apparently, this notion of

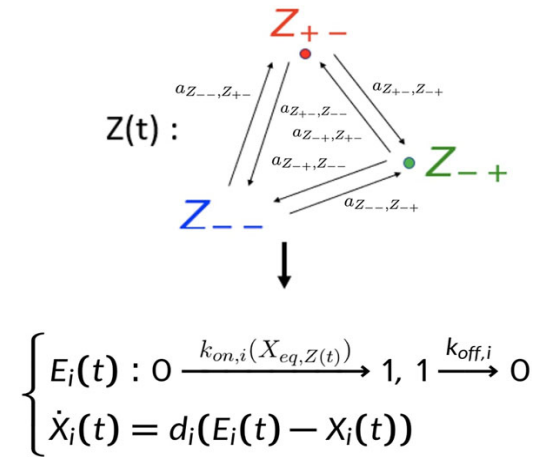


Fig. 11 Weak noise approximate model. The Markov chain on the set of basins Z is here illustrated by the one corresponding to the toggle-switch network of Fig. 3a

energy is not related to the one described previously. Once again, the difficulty for linking these two notions of energy comes from the fact that the dynamics associated to the "momentum" $p = \nabla_x V$, which is characterized by the velocity field v defined by the formula (23), is not the same that the protein dynamics associated in the weak noise limit to the marginal distribution on proteins $e^{-\frac{V(\cdot)}{\varepsilon}}$, which is defined by the promoters frequency v_{α_p}.

7.2 Mixture model

The results of Sect. 6 lead us to consider the coarse-grained model as promising for capturing the dynamics of the metastable system, even for realistic ε. We are now going to introduce a mixture model which provides an heuristic link between the protein dynamics and the coarse-grained model, and appears then promising for combining both simplicity and ability to describe the main ingredients of cell differentiation process.

When ε is small, a cell within a basin $Z_j \in Z$ is supposed to be most of the time close to its attractor: a rough approximation consists in identifying the activation rate of a promoter e_i in each basin by the dominant rate within the basin, corresponding to the value of $k_{on,i}$ on the attractor. For any gene $i = 1, \ldots, n$ and any basin $Z_j \in Z$, we can then consider:

$$\forall x \in Z_j : k_{on,i}(x) \approx k_{on,i}(X_{eq,Z_j}).$$

Combining this approximation of the functions $k_{on,i}$ by their main mode within each basin with the description of metastability provided in Sect. 3.2, we build another process described by the $2n + 1$-dimensional vector of variables $(Z(t), E(t), X(t))$, representing respectively the cell type, the promoter state and the protein concentration of all the genes (see Fig. 11).

Considering that the PDMP system spends in each basin a time long enough to equilibrate inside, we decide to approximate the distribution of the vector $(E(t), X(t))$ in a basin Z_j by its quasistationary distribution. It is then equivalent to the stationary distribution of a simple two states model with constant activation function, which is a

product of Beta distributions (Herbach et al. 2017). Thus, the marginal on proteins of the stationary distribution of this new model, that we denote u, can be approximated by a mixture of Beta distributions:

$$u \approx \sum_{Z_j \in Z} \mu_z(Z_j) \prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_j})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right), \tag{33}$$

where μ_z is the stationary distribution of the Markov chain characterizing the coarse-grained model.

In that point of view, the marginal distribution on proteins of a single cell X is characterized by a hidden Markov model: in each basin Z_j, which corresponds to the hidden variable, the vector X is randomly chosen under the quasistationary distribution u_{Z_j} of the reduced process $(E, X \mid Z_j)$. This simplified model provides a useful analytical link between the proteins distribution of the PDMP system (depending on the whole GRN) and the coarse-grained model parameters.

This mixture also provides an approximation for the potential of the system on Ω:

$$V(x) \approx -\ln\left(\sum_{Z_j \in Z} \mu_z(Z_j) \prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_j})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x)\right). \tag{34}$$

We remark that this new phenomenological model is a generalization of the local approximations of both the potential and the distribution within each basin that we have used for building the isocomittor surfaces and the score function of the AMS algorithm in "Appendices F.3 and G".

7.3 One application for the mixture model

An interesting application for the mixture approximation presented in Sect. 7.2 is the computation of the potential energy of the system, as defined in the previous section. The potential energy of a population of cells C located on $(x_c)_{c\in C}$ can be approximated by the sum $\sum_{c\in C} V(x_c)$, where V is defined by (34)

We represent in Fig. 12 the evolution of the potential energy of a population of cells during the differentiation process, simulated from the PDMP system associated to the toggle-switch network presented in "Appendix E". The population is initially centered on the attractor of the undifferentiated state Z_{--}. We observe that the potential energy reaches a peak before decreasing.

We remark that in Gao et al. (2020), the authors have revealed the universality of such feature during cell differentiation, for what they called the transcriptional uncertainty landscape, for many available single-cell gene expression data sets. This transcriptional uncertainty actually corresponds to the stationary potential V of our model, approximated for each cell from the exact stationary distribution of an uncoupled system of PDMPs (i.e with a diagonal interaction matrix). Although it cannot be formally linked to intracellular energetic spending yet, we can note that one of

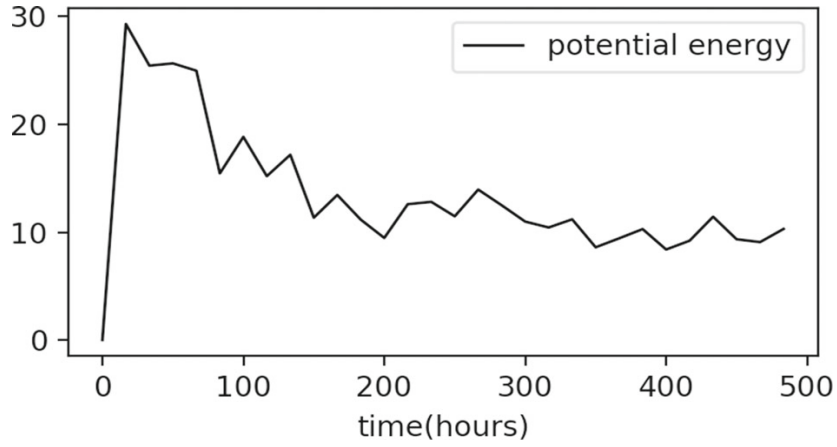


Fig. 12 Evolution of the potential energy V of a population of 500 cells along the differentiation process

the authors recently described a peak in energy consumption during the erythroid differentiation sequence (Richard et al. 2019).

The mixture model also paves the way for interpreting non-stationary behaviours. Indeed, let us denote $\mu_{z,t}$ the distribution of the basins at any time t. The mixture distribution can be used as a proxy for non stationary distributions of a PDMP system:

$$p_t \approx \sum_{Z_j \in Z} \mu_{z,t}(Z_j) \prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_j})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right).$$

In that case, the only time-dependent parameters are the coordinates of the vector $\mu_{z,t} \in [0,1]^m$ where m is the number of basins, and $\mu_{z,t} = \mu_z$ if t is such that the stationary distribution is reached. The parameters $(\mu_{z,t}(Z_j), \frac{k_{on,i}(X_{eq,Z_j})}{d_i}, \frac{k_{off,i}}{d_i})_{Z_j \in Z}$ could be inferred from omics data at any time t, for example with an EM algorithm (Pearce et al. 2019; Ma and Leijon 2009).

8 Conclusion

Reducing a model of gene expression to a discrete coarse-grained model is not a new challenge, (Lv et al. 2014; Lin and Galla 2016), and it is often hard to perform when the dimension is large. This reduction is closely linked to the notion of landscape through the quasipotential, the analysis of which has been often performed for non mechanistic models, where the random effects are considered as simple noise (Brackston et al. 2018; Wang et al. 2010), or for a number of genes limited to 2.

In this work, we propose a numerical method for approximating the transition rates of a multidimensional PDMP system modeling genes expression in a single cell. This method allows to compute these transition rates from the probabilities of some rare events, for which we have adapted an AMS algorithm. Although this method theoretically works for any GRN, the computation cost of the AMS algorithm may explode when both the number of genes increases and the scaling factor ε decreases.

In order to approximate these probabilities within the Large deviations context, we provided an explicit expression for the Hamiltonian and Lagrangian of a multidimensional PDMP system, we defined the Hamilton–Jacobi equation which characterizes the quasipotential, for any number of genes, and we provided the explicit expression of the associated variational problem which characterizes the landscape. We have deduced for some networks an analytical expression of the energetic costs of switching between the cell types, from which the transition rates can be computed. These approximations are accurate for a two-dimensional toggle-switch. We also verified that these analytical approximations seem accurate even for networks of 3 or 4 genes for which the energetic cost provided by the method is not proved to be optimal. However, testing the accuracy of this method for describing more complex networks would imply to build an approximate solution to the stationary Hamilton–Jacobi Eq. (25), which would be the subject of future works.

Finally, we have derived from the coarse-grained model a Beta-mixture model able to approximate the stationary behavior of a cell in the gene expression space. As far as we know, this is the first time that such an explicit link between a PDMP system describing cell differentiation and a non-Gaussian mixture model is proposed.

Altogether this work establishes a formal basis for the definition of a genetic/epigenetic landscape, given a GRN. It is tempting to now use the same formalism to assess the inverse problem of inferring the most likely GRN, given an (experimentally-determined) cell distribution in the gene expression space, a notoriously difficult task (Pratapa et al. 2020; Herbach et al. 2017).

Such random transitions between cell states have been recently proposed as the basis for facilitating the concomitant maintenance of transcriptional plasticity and stem cell robustness (Wheat et al. 2020). In this case, the authors have proposed a phenomenological view of the transition dynamics between states. Our work lays the foundation for formally connecting this cellular plasticity to the underlying GRN dynamics.

Finally our work provides the formal basis for the quantitative modelling of stochastic state transitions underlying the generation of diversity in cancer cells (Zhou et al. 2014; Gupta et al. 2011), including the generation of cancer stem cells (Tong et al. 2018).

Acknowledgements This work was supported by funding from French agency ANR (SingleStatOmics; ANR-18-CE45-0023-03). We thank Ulysse Herbach for having highlighted the notions of main modes for the stochastic hybrid model of gene expression, and for critical reading of the manuscript. We would like to thank the referees and the associated editor for carefully reading our manuscript and for their constructive comments which helped improving the quality of the paper. We also thank all members of the SBDM and Dracula teams, and of the SingleStatOmics project, for enlightening discussions. We also thank the BioSyL Federation and the LabEx Ecofect (ANR-11-LABX-0048) of the University of Lyon for inspiring scientific events.

Declarations

Code availability The code for reproducing the main figures of the article is available at https://gitbio.ens-lyon.fr/eventr01/jomb_reduction. It also contains the functions for the AMS algorithm, which is detailed in the appendix.

A Mechanistic model and fast transcription reduction

We recall briefly the full PDMP model, which is described in details in Herbach et al. (2017), based on a hybrid version of the well-established two-state model of gene expression (Ko 1991; Peccoud and Ycart 1995) including both mRNA and protein production (Shahrezaei and Swain 2008) and illustrated in Fig. 13.

A gene is described by the state of a promoter, which can be $\{on, off\}$. If the promoter is on, mRNAs will be transcripted with a rate s_m and degraded with a rate d_m. If it is off, only mRNA degradation occurs. Translation of mRNAs into proteins happens regardless of the promoter state at a rate s_p, and protein degradation at a rate d_m. Neglecting the molecular noise of proteins and mRNAs, we obtain the hybrid model:

$$\begin{cases} E(t) : 0 \xrightarrow{k_{on}} 1, 1 \xrightarrow{k_{off}} 0, \\ M'(t) = s_m E(t) - d_m M(t), \\ P'(t) = s_p M(t) - d_p P(t). \end{cases}$$

where $(E(t), M(t), P(t))$ denote respectively the promoter, mRNA and protein concentration at time t. As detailed in Sect. 2, the key idea is then to put this two-states model into a network by characterizing the jump rates of each gene by two specific functions $k_{on,i}$ and $k_{off,i}$, depending at any time on the protein vector $X(t)$.

In order to obtain the PDMP system (1) that we use throughout this article, we exploit the two modifications that are performed in Herbach et al. (2017) to this mechanistic model. First, the parameters s_m and s_p can be removed to obtain a dimensionless model, from which physical trajectories can be retrieved with a simple rescaling.

Second, a scaling analysis leads to simplify the model. Indeed, degradation rates play a crucial role in the dynamics of the system. The ratio $\frac{d_{m,i}}{d_{p,i}}$ controls the buffering of promoter noise by mRNAs and, since $k_{off,i} \gg k_{on,i}$, the ratio $\frac{k_{on,i}}{d_{m,i}}$ controls the buffering of mRNA noise by proteins. In line with several experiments (Albayrak et al. 2016; Li and Xie 2011), we consider that mRNA bursts are fast in regard to protein dynamics, i.e $\frac{d_{m,i}}{d_{p,i}} \gg 1$ with $\frac{k_{on,i}}{d_{m,i}}$ fixed. The correlation between mRNAs and proteins produced by the gene is then very small, and the model can be reduced by removing mRNA and making proteins directly depend on the promoters. We then obtain the PDMP system (1).

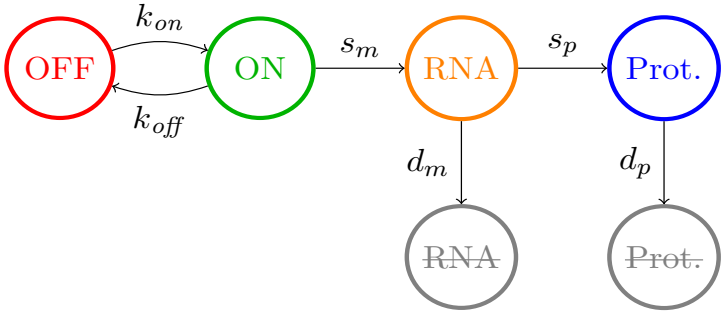


Fig. 13 The two-states model of gene expression (Herbach et al. 2017; Peccoud and Ycart 1995)

Denoting $\overline{k}_i$ the mean value of the function $k_{on,i}$, i.e its value where there is no interaction between gene i and the other genes, the value of a scaling factor $\varepsilon_i = \frac{d_{p,i}}{\overline{k}_i}$ can then be decomposed in two factors: one describing the ratio between the degradation rates of mRNA and proteins, $\frac{d_{p,i}}{d_{m,i}}$, which is evaluated around $1/5$ in Schwanhäusser et al. (2011), and one characterizing the ratio between promoter jumps frequency and the degradation rates of mRNA, $\frac{d_{m,i}}{\overline{k}_i}$. This last ratio is very difficult to estimate in practice. Assuming that it is smaller than 1, i.e that the mean exponential decay of mRNA when the promoter E_i is *off* is smaller than the mean activation rate, we can consider that ε_i is smaller than $1/5$. Finally, for obtaining the model (1), we consider two typical timescales $\bar{d}$ and $\bar{k}$, for the rates of proteins degradation and promoters activation respectively, such that for all genes i, $\frac{\overline{k}_i}{\bar{k}}$ and $\frac{d_{p,i}}{\bar{d}}$ are of order 1 (when the disparity between genes is not too important). We then define $\varepsilon = \frac{\bar{d}}{\bar{k}}$.

B Tensorial expression of the master equation of the PDMP system

We detail the tensorial expression of the master Eq. (2) for a two-dimensional network. We fix $\varepsilon = 1$ for the sake of simplicity.

The general form for the infinitesimal operator can be written:

$$Lu(t,e,x) = \langle F(e,x), \nabla_x u(t,e,x)\rangle + \sum_{e' \in P_E} Q(e,e')(x)u(t,e',x)$$

where F is the vectorial flow associated to the PDMP and Q the matrix associated to the jump operator.

A jump between two promoters states e, e' is possible only if there is exactly one gene for which the promoter has a different state in e than in e': in this case, we denote $e \sim e'$.

We have, for any x: $F(e,x) = (d_0(e_0 - x_0), \ldots, d_n(e_n - x_n))^T$. Then, for all $e \in P_E$, the infinitesimal operator can be written:

$$\begin{aligned} Lu(t,e,x) = &\sum_{i=1}^{n} F_i(e,x)\partial_{x_i} u(t,e,x) \\ &+ \sum_{\{e' | e' \sim e\}} \left(k_{on,i}(x)\delta_{e_i=0} + k_{off,i}\delta_{e_i=1}\right)\left(u(t,e',x) - u(t,e,x)\right). \end{aligned}$$

For a two-dimensional process ($n = 2$), there are four possible configurations for the promoter state: $e_{00} = (0,0)$, $e_{01} = (0,1)$, $e_{10} = (1,0)$, $e_{11} = (1,1)$. It is impossible to jump between the states e_{00} and e_{11}. If we denote $u(t,x)$ the four-dimensional vector: $(u_e(t,x))_{e \in P_E}$, we can write the infinitesimal operator in a matrix

form:

$$
\begin{aligned}
Lu(t,x) = &\underbrace{\begin{pmatrix} -d_1x_1 & 0 & 0 & 0 \\ 0 & -d_1x_1 & 0 & 0 \\ 0 & 0 & d_1(1-x_1) & 0 \\ 0 & 0 & 0 & d_1(1-x_1) \end{pmatrix}}_{F_1(x)} \begin{pmatrix} \partial_{x_1}u_{e_{00}}(t,x) \\ \partial_{x_1}u_{e_{01}}(t,x) \\ \partial_{x_1}u_{e_{10}}(t,x) \\ \partial_{x_1}u_{e_{11}}(t,x) \end{pmatrix} + \\
&\underbrace{\begin{pmatrix} -d_2x_2 & 0 & 0 & 0 \\ 0 & d_2(1-x_2) & 0 & 0 \\ 0 & 0 & -d_2x_2 & 0 \\ 0 & 0 & 0 & d_2(1-x_2) \end{pmatrix}}_{F_2(x)} \begin{pmatrix} \partial_{x_2}u_{e_{00}}(t,x) \\ \partial_{x_2}u_{e_{01}}(t,x) \\ \partial_{x_2}u_{e_{10}}(t,x) \\ \partial_{x_2}u_{e_{11}}(t,x) \end{pmatrix} + \\
&\underbrace{\begin{pmatrix} -k_{on,1}(x) & 0 & k_{on,1}(x) & 0 \\ 0 & -k_{on,1}(x) & 0 & k_{on,1}(x) \\ k_{off,1} & 0 & -k_{off,1} & 0 \\ 0 & k_{off,1} & 0 & -k_{off,1} \end{pmatrix}}_{Q_1(x)} \begin{pmatrix} u_{e_{00}}(t,x) \\ u_{e_{01}}(t,x) \\ u_{e_{10}}(t,x) \\ u_{e_{11}}(t,x) \end{pmatrix} + \\
&\underbrace{\begin{pmatrix} -k_{on,2}(x) & k_{on,2}(x) & 0 & 0 \\ k_{off,2} & -k_{off,2} & 0 & 0 \\ 0 & 0 & -k_{on,2}(x) & k_{on,2}(x) \\ 0 & 0 & k_{off,2} & -k_{off,2} \end{pmatrix}}_{Q_2(x)} \begin{pmatrix} u_{e_{00}}(t,x) \\ u_{e_{01}}(t,x) \\ u_{e_{10}}(t,x) \\ u_{e_{11}}(t,x) \end{pmatrix}.
\end{aligned}
$$

We remark that each of these matrices can be written as a tensorial product of the corresponding two-dimensional operator with the identity matrix:

- $F_1(x) = F^{(1)}(x) \otimes I_2$
- $F_2(x) = I_2 \otimes F^{(2)}(x)$
- $F^{(i)}(x) = \begin{pmatrix} -d_i x_i & 0 \\ 0 & d_i(1-x_i) \end{pmatrix}$
- $Q_1(x) = Q^{(1)}(x) \otimes I_2$
- $Q_2(x) = I_2 \otimes Q^{(2)}(x)$
- $Q^{(i)}(x) = \begin{pmatrix} -k_{on,i}(x) & k_{on,i}(x) \\ k_{off,i} & -k_{off,i} \end{pmatrix}$.

The master Eq. (2) is obtained by taking the adjoint operator of L:

$$
\frac{\partial u}{\partial t}(t,x) = L^*u(t,x) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(F_i u)(t,x) + \sum_{i=1}^{n} K_i u(t,x)
$$

where $K(x) = Q^T(x)$ is the transpose matrix of Q.

C Diffusion approximation

C.1 Definition of the SDE

In this section, we apply a key result of Pakdaman et al. (2012) to build the diffusion limit of the PDMP system (1). Let us denote $\overline{X}_t$ a trajectory satisfying the ODE system:

$$\dot{x}(t) = \overline{v}(x(t)),$$

where $\overline{v} : x \to d\left(\frac{k_{on}(x)}{k_{on}(x)+k_{off}} - x\right)$ characterizes the deterministic system (4). We consider the process Z_t^ε defined by:

$$Z_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}(X_t^\varepsilon - \overline{X}_t),$$

where $X_t\varepsilon$ verifies the PDMP system. Then, from the theorem 2.3 of Pakdaman et al. (2012) the sequence of processes $\{Z_t^\varepsilon\}_\varepsilon$ converges in law when $\varepsilon \to 0$ to a diffusion process which verifies the system:

$$dZ_t = \partial_x \overline{v}(\overline{X}_t)Z_t dt + \sigma(\overline{X}_t)dB_t, \tag{35}$$

where B_t denotes the Brownian motion. The diffusion matrix $\Sigma(x) = \sigma(x)\sigma^T(x)$ is defined by:

$$\forall i, j = 1, \ldots, n,\ \Sigma_{i,j}(x) := \sum_e 2W_i(x,e)\phi_j(x,e)\zeta(x,e),$$

where $\forall e \in P_E,\ W(x,e) = d(e-x) - \overline{v}(x)$, and ϕ is solution of a Poisson equation:

$$\begin{cases} \forall e \in P_E, \forall i = 1, \ldots, n : \sum_{e'} Q_{ee'}(x)\phi_i(x,e') = -W_i(x,e), \\ \sum\limits_{e \in P_E} \phi_i(x,e)\zeta(x,e) = 0. \end{cases} \tag{36}$$

Let ζ be a probability vector in S_E representing the stationary measure of the jump process on promoters knowing proteins : $\forall x \in \Omega,\ \zeta(x,\cdot)Q(x) = 0$. We have: $\forall e \in P_E, \zeta(x,e) = \prod\limits_{i=1}^{n} \frac{k_{on,i}^{e_i}(x)k_{off,i}^{1-e_i}}{k_{on,i}(x)+k_{off,i}}$.

It is straightforward to see that for all $i = 1, \ldots, n$: $W_i(x,e) = d_i\left(e_i - \frac{k_{on,i}(x)}{k_{on,i}(x)+k_{off,i}}\right)$. Then, let us define ϕ such that:

$$\forall e \in P_E, \forall i = 1, \ldots, n : \phi_i(x,e) = \frac{d_i}{k_{on,i}(x) + k_{off,i}}\left(e_i - \frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}}\right).$$

We verify that this vector ϕ is solution to the Poisson equation (36) for all x. The matrix $\Sigma(x)$ is then a diagonal matrix defined by:

$$\begin{aligned}\forall i = 1, \ldots, n : \ \Sigma_{ii}(x) &= 2\frac{d_i^2 k_{on,i}^2(x)k_{off,i} + d_i^2 k_{off,i}^2 k_{on,i}(x)}{(k_{on,i}(x) + k_{off,i})^4} \\ &= 2\frac{d_i^2 k_{on,i}(x)k_{off,i}}{(k_{on,i}(x) + k_{off,i})^3}. \end{aligned} \tag{37}$$

For all $x \in \Omega$, the matrix $\sigma(x)$ is then also diagonal and defined by:

$$\forall i = 1, \ldots, n, \ \sigma_{ii}(x) = \sqrt{\frac{2d_i^2 k_{on,i}(x)k_{off,i}}{(k_{on,i}(x) + k_{off,i})^3}},$$

and we have defined all the terms of the diffusion limit (35).

C.2 The Lagrangian of the diffusion approximation is a second-order approximation of the Lagrangian of the PDMP system

It is well known that the diffusion approximation satisfies a LDP of the form (12) (Freidlin and Wentzell 2012). The formula (37) allows to define the Lagrangian associated to this LDP, that we denote L_d. From the theorem 2.1 of Freidlin and Wentzell (2012), we have:

$$\begin{aligned}&\forall x, v \in \Omega \times \Omega^v(x) : \ L_d(x, v) \\ &= \sum_{i=1}^{n} \frac{(k_{on,i}(x) + k_{off,i})^3}{4d_i^2 k_{on,i}(x)k_{off,i}} \left(v_i - d_i \left(\frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} - x_i \right) \right)^2. \end{aligned} \tag{38}$$

Note that for any fixed $x \in \Omega$, $L_d(x, \cdot)$ is a quadratic function.

We recall that the Lagrangian associated to the LDP for the PDMP system, that we found in Theorem 3, is defined for all $x, v \in \Omega \times \Omega^v(x)$ by:

$$L(x, v) = \sum_{i=1}^{n} \left(\sqrt{k_{off,i} \frac{v_i + d_i x_i}{d_i}} - \sqrt{k_{on,i}(x) \frac{d_i(1 - x_i) - v_i}{d_i}} \right)^2.$$

Expanding this Lagrangian with respect to v around $\overline{v}$ (the drift of the relaxation trajectories), we obtain:

$$L(x, v) = \sum_{i=1}^{n} \left(\frac{k_{on,i}(x) + k_{off,i}}{2d_i \sqrt{\frac{k_{on,i}(x)k_{off,i}}{k_{on,i}(x)+k_{off,i}}}} \right)^2 (v_i - \overline{v_i})^2 + o\left(v_i - \overline{v_i}\right)^2 = L_d(x, v) + o\left(v_i - \overline{v_i}\right)^2.$$

Thus, we proved that the Lagrangian of the diffusion approximation of the PDMP process corresponds to the two first order terms in $(v_i - \overline{v_i})$ of the Taylor expansion of the real Lagrangian.

D Example of interaction function

We recall that we assume that the vector k_{off} does not depend on the protein vector.

The specific interaction function chosen comes from a model of the molecular interactions at the promoter level, described in Herbach et al. (2017):

$$k_{on,i}(X) = \frac{k_{0,i} + k_{1,i}(\sigma_i X_i)^{m_{ii}} \Phi_i(X)}{1 + (\sigma_i X_i)^{m_{ii}} \Phi_i(X)}, \tag{39}$$

with:

- $k_{0,i}$ the basal rate of expression of gene i,
- $k_{1,i}$ the maximal rate of expression of gene i,
- $m_{i,j}$ an interaction exponent, representing the power of the interaction between genes i and j,
- σ_i is the rescaling factor depending on the parameters of the full model including mRNAs,
- θ a matrix defining the interactions between genes, corresponding to a matrix with diagonal terms defining external stimuli, and
- $\Phi_i(X) = e^{\theta_{i,i}} \prod_{j \neq i} \frac{1+e^{\theta_{j,i}+\theta_{j,j}(\sigma_j X_j)^{m_{ji}}}}{1+e^{\theta_{j,j}(\sigma_j X_j)^{m_{ji}}}}$.

For a two symmetric two-dimensional network, we have for any $x = (x_1, x_2) \in \Omega$:

$$\frac{\partial_{x_2} k_{on,1}(x)}{x_1} = \frac{m_{21} e^{\theta_{22}} x_2^{m21-1} e^{\theta_{11}} x_1^{m11-1} (1 - e^{\theta_{12}})}{1 + e^{\theta_{22}} x_2^{m21} + e^{\theta_{11}} x_1^{m11} + x_2^{m21} x_1^{m11} e^{\theta_{11}+\theta_{22}+\theta_{12}}}.$$

When $m_{11} = m_{22} = m_{12} = m_{21}$ and $\theta_{12} = \theta_{21}$, we have then for every $x \in \Omega$:

$$\frac{\partial_{x_2} k_{on,1}(x)}{x_1} = \frac{\partial_{x_1} k_{on,2}(x)}{x_2}.$$

Thus, for all $x \in \Omega$, when $d_1 = d_2$ we have:

$$\partial_{x_2} \left(-\frac{k_{on,1}(x)}{d_1 x_1} + \frac{k_{off,1}}{d_1(1 - x_1)} \right) = \partial_{x_1} \left(-\frac{k_{on,2}(x)}{d_2 x_2} + \frac{k_{off,2}}{d_2(1 - x_2)} \right).$$

As a consequence, owing to the Poincaré lemma, there exists a function $V \in C^1(\Omega, \mathbb{R})$ such that the condition (26) is satisfied: one has

$$\forall i = 1, 2 : \partial_{x_i} V(x) = -\frac{k_{on,i}(x)}{d_i x_i} + \frac{k_{off,i}}{d_i(1 - x_i)}.$$

E Description of the toggle-switch network

This table describes the parameters of the symmetric two-dimensional toggle-switch used all along the article. These values correspond to the parameters used for the simulations. The rescaling in time by the parameter scale $\bar{d}$, for the model presented in Sect. 2, corresponds to divide every $k_{0,i}$, $k_{1,i}$, d_i by $\bar{d} = 0.2$. The mean values $\bar{k}_i$ and d_i are then, as expected, of order 1 for every gene i.

(i, j)	$k_{0,i}$	$k_{1,i}$	d_i	σ_i	$m_{i,i}$	$m_{i,j}$	$\theta_{i,i}$	$\theta_{i,j}$	$k_{off,i}$
(1,2)	$0,012/\varepsilon$	$0.39/\varepsilon$	0, 2	5	3	3	7	−7	$1,25/\varepsilon$
(2,1)	$0,012/\varepsilon$	$0.39/\varepsilon$	0, 2	5	3	3	7	−7	$1,25/\varepsilon$

F Details on the approximation of the transition rate as a function of probability (7)

In this section, we adapt the method developed in Cérou et al. (2011) to justify the formula (8) provided in Sect. 4.1, which approximate for every pair of basins (Z_i, Z_j) the transition rate a_{ij} as a function of the probability (7).

F.1 General setting

Let us consider r, R such that $0 < r < R$, we recall that γ_{Z_i} and Γ_{Z_i} denote respectively the r-neighborhood and the R-neighborhood of the attractor X_{eq,Z_i}. Let us consider a random path X_t^ε of the PDMP system, with initial condition $X_0^\varepsilon = x_0 \in \partial\Gamma_{Z_i}$. We define the series of stopping times $(\mu_l^\varepsilon)_{l\in\mathbb{N}}$, $(\sigma_l^\varepsilon)_{l\in\mathbb{N}^*}$ such that $\mu_0^\varepsilon = 0$ and for all $l \in \mathbb{N}^*$:

- $\sigma_l^\varepsilon = \inf\{t \geq \mu_{l-1}^\varepsilon \mid X_t^\varepsilon \in \{\gamma_{Z_i} \cup \{\bigcup_{k\neq i} Z_k\}\}$,
- $\mu_l^\varepsilon = \inf\{t \geq \sigma_l^\varepsilon \mid X_t^\varepsilon \in Z_i \setminus \Gamma_{Z_i}\}$.

We then define $Y_l^\varepsilon = X_{\sigma_l}^\varepsilon$. If $Y_l^\varepsilon \in Z_j$, we set $\forall k > l : \sigma_k^\varepsilon = \mu_k^\varepsilon = \infty$ and the chain Y_l^ε stops.

From the formula (6) characterizing the transition rates, we can write:

$$a_{ij}^\varepsilon \simeq \frac{\mathbb{P}_{x_0}(\hat{Z}_1^\varepsilon = Z_j)}{\mathbb{E}_{x_0}(\tau_1^\varepsilon)} = \frac{\mathbb{P}_{x_0}\left(T_{Z_j}^\varepsilon < T_{\{Z\setminus\{\overline{Z_i}\cup\overline{Z_j}\}\}}^\varepsilon\right)}{\mathbb{E}(\sigma_{w_i^\varepsilon})}, \tag{40}$$

where we define the random variable: $w_i^\varepsilon = \inf\{l \mid Y_l^\varepsilon \in \bigcup_{k\neq i} Z_k\}$.

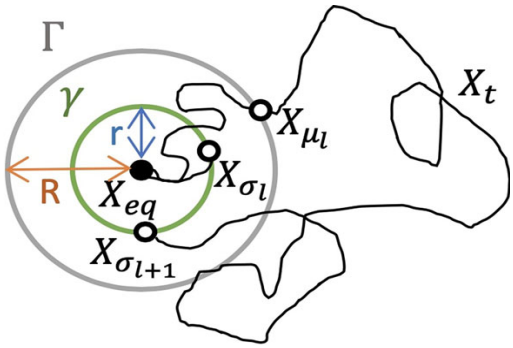


Fig. 14 Illustration of the stopping times σ_l and μ_l describing respectively the l^{th} entrance of a random path X_t in a r-neighborhood γ of an attractor X_{eq}, and its l^{th} exit from a R-neighborhood Γ

Let us denote $\overline{T}^{\varepsilon}_{Z_i,Z_i} = \mathbb{E}(\sigma^{\varepsilon}_{l+1} - \sigma^{\varepsilon}_{l} \mid Y^{\varepsilon}_{l} \in \gamma_{Z_i},\ Y^{\varepsilon}_{l+1} \in \gamma_{Z_i})$. We can make the following approximation:

$$\mathbb{E}(\sigma_{w^{\varepsilon}_i}) \simeq \mathbb{E}(w^{\varepsilon}_i) \times \overline{T}^{\varepsilon}_{Z_i,Z_i}.$$

Indeed, the quantity on the left hand side is close to the mean number of attempts for reaching, from $\partial\Gamma_{Z_i}$, a basin Z_k, $k \neq i$, before γ_{Z_i}, which is equal to $\mathbb{E}(w^{\varepsilon}_i)$, multiplied by the mean time of each attempt (knowing that at each step l, $Y^{\varepsilon}_{l} \in \gamma_{Z_i}$), which is exactly $\overline{T}^{\varepsilon}_{Z_i,Z_i}$. We should add the mean time for reaching ∂Z_j from $\partial\Gamma_{Z_i}$ at the last step, but it is negligible when the number of attempts is large, which is the case in the small noise limit.

F.2 Method when $\partial\Gamma_{Z_i}$ is reduced to a single point

We consider the case when $\partial\Gamma_{Z_i}$ is reduced to a single point. It can happen for example when we consider only one gene ($\Omega = (0, 1)$) and when the attractor X_{eq,Z_i} is located at a distance smaller than r from one of the boundaries of the gene expression space ($X_{eq,Z_i} < r$ or $X_{eq,Z_i} > 1 - r$). In such situation, a random path crosses necessarily the same point x_0 to both exit Γ_{Z_i} and come back to γ_{Z_i} (if it does not reach a basin Z_j before): the Markov property of the PDMP process then justifies that the quantities $\mathbb{E}\left(\sigma^{\varepsilon}_{l} - \mu^{\varepsilon}_{l-1} \mid Y^{\varepsilon}_{l-1} \notin Z_j\right)$ and $\mathbb{E}\left(\mu^{\varepsilon}_{l} - \sigma^{\varepsilon}_{l} \mid Y^{\varepsilon}_{l} \notin Z_j\right)$ do not depend of l. Then $\mathbb{1}_{Y^{\varepsilon}_{l} \in Z_j}$ behaves like a discrete homogeneous Markov chain with two states, 1 being absorbing.

Let us define a second random variable $W^{\varepsilon}_{ij} = \inf\{l \mid Y^{\varepsilon}_{l} \in Z_j\}$. The homogeneity of the Markov chain $\mathbb{1}_{Y^{\varepsilon}_{l} \in Z_j}$ ensures that W^{ε}_{ij} follows a geometric distribution. Its expected value is then the reverse of the parameter of the geometric law, i.e: $\mathbb{E}(W^{\varepsilon}_{ij}) = (p^{\varepsilon}_{ij}(x_0))^{-1}$.

Moreover, it is straightforward to see that from the same reasoning applied to any Z_k, $k \neq i$:

$$\mathbb{E}(w^{\varepsilon}_i) = \frac{1}{\sum\limits_{k\neq i} p^{\varepsilon}_{ik}(x_0)} = \frac{1}{p^{\varepsilon}_{ij}(x_0)} \frac{p^{\varepsilon}_{ij}(x_0)}{\sum\limits_{k\neq i} p^{\varepsilon}_{ik}(x_0)} = \mathbb{E}(W^{\varepsilon}_{ij}) \times \mathbb{P}_{x_0}\left(T^{\varepsilon}_{Z_j} < T^{\varepsilon}_{\{Z\setminus\{\overline{Z_i}\cup\overline{Z_j}\}\}}\right).$$

Thus, from (40) we can approximate the transition rate by the formula (8).

F.3 Method in the general case

The difficulty for generalizing the approach described above, when $\partial\Gamma_{Z_i}$ is not reduced to a single point, is to keep the Markov property, which has been used to cut the trajectories into pieces. Heuristically, the same argument which led us to approximate the PDMP system by a Markov jump process can be used to justify the asymptotic independence on l of the quantity $\mathbb{E}(\mu_l^\varepsilon - \sigma_l^\varepsilon \mid Y_l^\varepsilon \notin Z_j)$: for $\varepsilon \ll 1$, any trajectory starting on $\partial\gamma_{Z_i}$ will rapidly loose the memory of its starting point after a mixing time within γ_{Z_i}. But it is more complicated to conclude on the independence from l of the quantity $\mathbb{E}(\sigma_l^\varepsilon - \mu_{l-1}^\varepsilon \mid Y_{l-1}^\varepsilon \notin Z_j)$, which may depend on the position of $X^\varepsilon_{\mu_{l-1}}$ when the gene expression space is multidimensional.

We introduce two hypersurfaces $\gamma_{min,Z_i} = \{x \in Z_i,\, p_{ij}^\varepsilon(x) = c_1\}$ and $\Gamma_{min,Z_i} = \{x \in Z_i,\, p_{ij}^\varepsilon(x) = c_2\}$, where $c_1 < c_2$ are two small constants. We substitute to the squared euclidean distance, used for characterizing the neighborhood γ_{Z_i} and Γ_{Z_i}, a new function based on the probability of reaching the (unknown) boundary: $\forall x, y \in Z_i,\ \| x - y \|^2 \leftarrow \mid p_{ij}^\varepsilon(x) - p_{ij}^\varepsilon(y) \mid$. The function p_{ij}^ε is generally called committor, and the hypersurfaces γ_{min,Z_i} and Γ_{min,Z_i} isocommittor surfaces. The committor function is not known in general; if it was, employing a Monte-Carlo method would not be necessary for obtaining the probabilities (7). However, it can be approximated from the potential of the PDMP system within each basin, defined in the equilibrium case by the well-known Boltzman law: $V = -\ln(\hat{u})$, $\hat{u}$ being the marginal on proteins of the stationary distribution of the process. Indeed, for reasons that are precisely the subject of Sect. 4.2 (studied within the context of Large deviations), the probability $p_{ij}^\varepsilon(x)$ is generally linked in the weak noise limit to the function V by the relation:

$$\forall x \in Z_i : \ p_{ij}^\varepsilon(x) \underset{\varepsilon \to 0}{\sim} C_{ij} e^{V(x)/\varepsilon},$$

where C_{ij} is a constant specific to each pair of basins (Z_i, Z_j). We remark that when ε is small, a cell within a basin $Z_j \in Z$ is supposed to be most of the time close to its attractor: a rough approximation could lead to identify the activation rate of a promoter e_i in each basin by the dominant rate within the basin, corresponding to the value of $k_{on,i}$ on the attractor. For any gene $i = 1, \ldots, n$ and any basin $Z_j \in Z$, we can then approximate:

$$\forall x \in Z_j : \ k_{on,i}(x) \approx k_{on,i}(X_{eq,Z_j}).$$

Under this assumption, the stationary distribution of the process is close to the stationary distribution of a simple two states model with constant activation function, which is a product of Beta distributions (Herbach et al. 2017). We then obtain an approximation of the marginal on proteins of the stationary distribution within each basin Z_j:

$$p_{Z_j} \approx \prod_{i=1}^{n} \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_j})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right),$$

By construction, this approximation is going to be better in a small neighborhood of the attractor X_{eq,Z_j}. Thus, this expression provides an approximation of the potential V around each attractor X_{eq,Z_j}:

$$V \approx -\ln\left(p_{Z_j}\right). \tag{41}$$

In every basin Z_i, and for all $x \in Z_i$ close to the attractor, the hypersurfaces where p_{ij}^{ε} is constant will be then well approximated by the hypersurfaces where the explicitly known function $\xi_{Z_i} = -\ln\left(p_{Z_i}\right)$ is constant.

For each attractor X_{eq,Z_i}, we can then approximate the two isocommittor surfaces described previously:

$$\begin{cases} \gamma_{min,Z_i} & \simeq \{x \in Z_i \mid \xi_{Z_i}(x) = c_1'\} \\ \Gamma_{min,Z_i} & \simeq \{x \in Z_i \mid \xi_{Z_i}(x) = c_2'\}, \end{cases} \tag{42}$$

where c_1' and c_2' are two constants such that $\xi_{Z_i}(X_{eq,Z_i}) < c_1' < c_2'$.

We then replace γ_{Z_i} and Γ_{Z_i} by, respectively, γ_{min,Z_i} and Γ_{min,Z_i} in the definitions of the stopping times μ_l^{ε} and σ_l^{ε} provided in Sect. F.1. From the proposition 1. of Cérou et al. (2011), we obtain that, as in the simple case described in Sect. F.2, $\mathbb{E}(\sigma_l^{\varepsilon} - \mu_l^{\varepsilon} \mid Y_l^{\varepsilon} \notin Z_j)$ is independent of l. Defining $W_{ij}^{\varepsilon} = \inf\{l \mid Y_l^{\varepsilon} \in Z_j\}$, the definition of Γ_{min,Z_i} allows to ensure that W_{ij}^{ε} does not depend on the point x_0 of Γ_{min,Z_i} which is crossed at each step $l' < W_{ij}^{\varepsilon}$. This random variable follows then a geometric distribution, with expected value $(p_{ij}^{\varepsilon}(x_0))^{-1}$, and we can derive an expression of the form (8).

G AMS algorithm

We use an Adaptive Multilevel Splitting algorithm (AMS) described in Bréhier et al. (2016). The algorithm provides for every Borel sets (A, B) an unbiased estimator of the probability:

$$\mathbb{P}_x^{\varepsilon}(\tau_A^{\varepsilon} < \tau_B^{\varepsilon}).$$

It is supposed that the random process attains easily A from x, more often than B, called the target set.

The crucial ingredient we need to introduce is a score function $\xi(\cdot)$ to quantify the adaptive levels describing how close we are from the target set B from any point x. The variance of the algorithm strongly depends on the choice of this function.

The optimal score function is the function $x \mapsto \mathbb{P}_x^{\varepsilon}(\tau_A^{\varepsilon} < \tau_B^{\varepsilon})$ itself, called the committor which is unknown. It is proved, at least for multilevel splitting algorithms applied to stochastic differential equations in Dean and Dupuis (2009), Budhiraja and Dupuis (2019), that if a certain scalar multiplied by the score function is solution of the associated stationary Hamilton–Jacobi equation, where the Hamiltonian comes from

the Large deviations setting, the number of iterations by the algorithm to estimate the probability in a fixed interval confidence grows sub-exponentially in ε.

For the problem studied in this article, for every basin $Z_j \in Z$, we want to estimates probabilities substituting A to γ_j and B to another basin Z_k, $k \neq j$. Using the approximation of V given by the expression (34), we obtain the following score function, up to a specific constant specific to each basin:

$$\xi(x) = -\ln\left(\sup_{Z_m \in Z}\left(\prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_m})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x)\right)\right).$$

We remark that this last approximation allows to retrieve the definition of the local potential (41) defined on Appendix F.3, when the boundary of the basins are approximated by the leading term in the Beta mixture. The approximation is justified by the fact that for small ε, the Beta distributions are very concentrated around their centers, meaning that for every basin $Z_k \in Z$, $k \neq j$:

$$\forall x \in Z_j,\ \prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_k})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x) \ll \prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_j})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x).$$

We supposed that $\forall Z_j \in Z$, $\mu_z(Z_j) > 0$, where μ_z denotes the distributions on the basins. This is a consequence of the more general assumption that the stationary distribution of the PDMP system is positive on the whole gene expression space, which is necessary for rigorously deriving an analogy of the Gibbs distribution for the PDMP system (see Sect. 4.2.2).

We modify the score function to be adapted for the study of the transitions from each basin Z_j to Z_k, $k \neq j$:

$$\begin{aligned}\xi_k(x) = &-\ln\left(\sup_{Z_m \in Z}\left(\prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_m})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x)\right)\right)\\ &+\ln\left(\prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_k})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x)\right).\end{aligned}$$

This function is specific to each transition to a basin Z_k but defined in the whole gene expression space. We verify: $\xi_k(x) \leq 0$ if $x \in \Omega \setminus Z_k$ and $\xi_k(x) = 0$ if $x \in Z_k$. We use ξ_k as the score function for the AMS algorithm.

In order to estimate probabilities of the type $\mathbb{P}_x^\varepsilon(\tau_{Z_k}^\varepsilon < \tau_{\gamma_j}^\varepsilon)$ for $x \in Z_j$, we need to approximate the boundaries of the basins of attraction, which are unknown. For this sake, we use again the approximate potential function $\xi \approx V$ to approximate the basins only from the knowledge of their attractor:

$$\forall Z_k \in Z :\ Z_k \simeq \{x \in \Omega \mid \underset{Z_m \in Z}{\text{argmax}}\left(\prod_i \text{Beta}\left(\frac{k_{on,i}(X_{eq,Z_m})}{\varepsilon d_i}, \frac{k_{off,i}}{\varepsilon d_i}\right)(x)\right) = Z_k\}.$$

We use the Adaptative Multilevel Splitting algorithm described in Section 4 of Bréhier et al. (2016), with two slight modifications in order to take into account the differences due to the underlying model and objectives:

• First, a random path associated to the PDMP system does not depend only on the protein state but is characterized at each time t by the 2n-dimensional vector: (X_t, E_t). For any simulated random path, we then need to associate an initial promoter state. However, we know that in the weak noise limit, for a protein state close to the attractor of a basin, the promoter states are rapidly going to be sampled by the quasistationary distribution: heuristically, this initial promoter state will not affect the algorithm. We decide to initially choose it randomly under the quasistationary distribution. For every $x_0 \in \Gamma_{min,j}$ beginning a random path in a basin Z_j, we choose for the promoter state of any gene i, e_{i_0}, following a Bernoulli distribution:

$$e_{i_0} \sim B\left(1, \frac{k_{on,i}(x_0)}{k_{off,i} + k_{on,i}(x_0)}\right).$$

• Compared with (Bréhier et al. 2016), an advanced algorithm is used to improve the sampling of the entrance time in a set $\gamma_{min,j}$. In practice timestepping is required to approximate the protein dynamics, and it may happen that the exact solution enters $\gamma_{min,j}$ between two time steps, whereas the discrete-time approximation remains outside $\gamma_{min,j}$. We propose a variant of the algorithm studied in Gobet (2000) for diffusion processes, where a Brownian Bridge approximation gives a more accurate way to test entrance in the set $\gamma_{min,j}$.

In the case of the PDMP system, we replace the Brownian Bridge approximation, by the solution of the ODE describing the protein dynamics: considering that the promoter state e remains constant between two timepoints, the protein concentration of every gene i, $x_i(t)$ is a solution of the ODE: $\dot{x_i}(t) = d_i(e_i - x_i(t))$, which implies:

$$\forall t \in [0, \Delta t] : \; x_i(t) = e_i + (x_i(0) - e_i)e^{-td_i}.$$

We show that for one gene, the problem can be easily solved. Indeed, let us denote $X_{i_{eq,Z_j}}$ the i^{th} component of the vector X_{eq,Z_j}. The function: $f_i(t) = (x_i(t) - X_{eq,Z_{j_i}})^2$ is differentiable and its derivative

$$f_i'(t) = -2d_i(x_i(0) - e_i)e^{-td_i}((e_i - X_{i_{eq,Z_j}}) + (x_i(0) - e_i)e^{-td_i}),$$

vanishes if and only if $(x_i(0) - e_i)e^{-td_i} = (X_{i_{eq,Z_j}} - e_i)$, i.e when

$$t = \frac{1}{d_i} \ln\left(\frac{X_{i_{eq,Z_j}} - e_i}{x_i(0) - e_i}\right) = c_i.$$

Then, if $c_i \leq 0$ or $c_i \geq \Delta t$, the minimum of the squared euclidean distance of the i-th coordinate of the path to the attractor is reached at one of the points $x_i(0)$ or $x_i(\Delta t)$. If $0 \leq c_i \leq \Delta t$, the extremum is reached at $x_i(c_i)$. This value, if it is a minimum, allows

us to determine if the process has reached any neighborhood of an attractor X_{eq,Z_j} between two timepoints.

For more than one gene, the minimum of the sum: $|| x - X_{eq,Z_j} ||^2 = \sum_{i=1}^{n} (x_i(t) - X_{i_{eq,Z_j}})^2$ is more complicated to find. If for all $i = 1, \ldots, n$, $d_i = d$, which is the case of the two-dimensional toggle-switch studied in Sect. 6, the extremum can be explicitly computed:

$$t = \frac{1}{d} \ln \left(\frac{\sum_{i=1}^{n} (X_{i_{eq,Z_j}} - e_i)(x_i(0) - e_i)}{\sum_{i=1}^{n} (x_i(0) - e_i)^2} \right) = c.$$

But we recall that for more than one gene, the set of interest is the isocommittor surface $\gamma_{min,j}$ and not a neighborhood γ_j. An approximation consists in identifying $\gamma_{min,j}$ to the r-neighborhood of X_{eq,Z_j}, where r is the mean value of $|| x - X_{eq,Z_j} ||^2$ for $x \in \gamma_{min,j}$.

If the parameters d_i are not all similar, we have to make the hypothesis that the minimum is close to the minimum for each gene. In this case, we just verify that for any gene i, the value of the minimum $x_i(c_i)$ for every gene is not in the set $\{x_i \mid x \in \gamma_{Z_j}\}$: if it is the case for one gene, we consider that the process has reached the neighborhood γ_{Z_j} of the basins Z_j between the two timepoints.

H Proofs of Theorems 4 and 5

First, we recall the theorem of characteristics applied to Hamilton–Jacobi equation (Evans 2010), which states that for every solution $V \in C^1(\Omega, \mathbb{R})$ of (25), the system (16) associated to V

$$\begin{cases} p(t) &= \nabla_x V(\phi(t)) \\ \dot{\phi}(t) &= \nabla_p H(\phi(t), p(t)), \end{cases}$$

is equivalent to the following system of ODEs on $(x, p) \in \Omega \times \mathbb{R}^n$, for $x(0) = \phi(0)$ and $p(0) = \nabla_x V(x(0))$:

$$\begin{cases} \dot{p}(t) &= -\nabla_x H\,(x(t), p(t)) \\ \dot{x}(t) &= \nabla_p H\,(x(t), p(t))\,. \end{cases}$$

A direct consequence of this equivalence with an ODE system is that two optimal trajectories associated to two solutions of the stationary Hamilton–Jacobi equation cannot cross each other with the same velocity. We then have the following lemma:

Lemma 2 *Let V_1 and V_2 be two solutions of (25) in $C^1(\Omega, \mathbb{R})$.*

For any trajectories $\phi_t^1, \phi_t^2 \in C_{0T}^{1,pw}(\Omega)$ *solutions of the system* (16) *associated respectively to* V_1 *and* V_2*, if there exists* $t \in [0, T]$ *such that* $\phi^1(t) = \phi^2(t)$ *and* $\dot{\phi}^1(t) = \dot{\phi}^2(t)$*, then one has* $\phi^1(t) = \phi^2(t)$ *for all* $t \in [0, T]$.

This corollary is important for the two first items of the proof of Theorem 4:

Corollary 1 *For any solution* $V \in C^1(\Omega, \mathbb{R})$ *of* (25) *and any trajectory* $\phi_t \in C_{0T}^{1,pw}(\Omega)$ *satisfying the system* (16) *associated to* V*, we have the equivalence:*

$$\exists\, t \in [0, T],\ \forall i \in \{1, \dots, n\} :\ \dot{\phi}_i(t) = d_i \left(\frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} - \phi_i(t) \right)$$
$$\iff \forall t \in [0, T] :\ \nabla_x V(\phi(t)) = 0.$$

Proof We recall that the relaxation trajectories correspond to trajectories satisfying the system (16) associated to a constant function V, i.e such that $\nabla_x V = 0$ on the whole trajectory. At any time t, the correspondence between any velocity field v of Ω^v and a unique vector field p, proved in Theorem 3 with the relation (24), allows to ensure that:

$$\forall i \in \{1, \dots, n\} :\ \dot{\phi}_i(t) = d_i \left(\frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} - \phi_i(t) \right) \iff p(t)$$
$$= \nabla_x V(\phi(t)) = 0.$$

The Lemma 2 ensures that any trajectory which verifies the same velocity field than a relaxation trajectory at a given time t is a relaxation trajectory: we can then conclude. □

Finally, the following lemma is important for the first item of the proof of Theorem 4:

Lemma 3 $\forall i \in \{1, \dots, n\},\ \forall x \in \Omega$ *we have:*

- $\dfrac{\partial}{\partial p_i} H(x, p') = 0 \iff H_i(x, p_i') = \min\limits_{p_i \in \mathbb{R}} H_i(x, p_i)$,
- $\min\limits_{p_i \in \mathbb{R}} H_i(x, p_i) \le 0$,
- $\min\limits_{p_i \in \mathbb{R}} H_i(x, p_i) = 0 \iff x_i = \dfrac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}}$.

Proof We have seen in the proof of Theorem 2 that for all $i = 1, \dots, n$ and for all $x \in \Omega$, $H_i(x, \cdot)$ is strictly convex, and that $H_i(x, p_i) \to \infty$ as $p_i \to \pm\infty$. Moreover, $H_i(x, p_i)$ vanishes on two points $p_{i_1} = 0$ and $p_{i_2} = -\frac{k_{on,i}(x)}{d_i x_i} + \frac{k_{off,i}(x)}{d_i(1-x_i)}$ inside $\mathbb{R}$.

Then, the min on p_i is reached on the unique critical point $p_i' \in [p_{i_1}, p_{i_2}]$, and we have: $H(x, p_i') = \min\limits_{p_i \in \mathbb{R}} H_i(x, p_i) \le 0$.

Finally:

$$\min_{p_i \in \mathbb{R}} H_i(x, p_i) = 0 \iff p_i' = p_{i_1} = p_{i_2} = 0$$

$$\iff x_i = \frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}}.$$

□

We now prove the theorem 4.

Proof of Theorem 4.(i) We consider a trajectory ϕ_t satisfying the system (16) associated to V. We recall that the Fenchel-Legendre expression of the Lagrangian allows to state that the vector field p associated to the velocity field $\dot{\phi}(t)$ by the relation (24) is precisely $p = \nabla_x V(\phi(t))$. When V is such that $H(\cdot, \nabla_x V(\cdot)) = 0$ on Ω, we have then for any time t:

$$L(\phi(t), \dot{\phi}(t)) = \sum_{i=1}^{n} \partial_{x_i} V(\phi(t)) \dot{\phi}_i(t). \tag{43}$$

We recall that from Theorem 3 :

$$L(\phi(t), \dot{\phi}(t)) = 0 \iff \left(\forall i = 1, \ldots, n : \dot{\phi}_i(t) = d_i \left(\frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} - \phi_i(t) \right) \right).$$

From this and (43), we deduce that for such an optimal trajectory:

$$\dot{\phi}(t) = 0 \implies \left(\forall i = 1, \ldots, n : \dot{\phi}_i(t) = d_i \left(\frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} - \phi_i(t) \right) \right).$$

The velocity field then vanishes only at the equilibrium points of the deterministic system.

Conversely, we recall that for such trajectory we have for any t:

$$\begin{cases} \dot{\phi}(t) = \frac{\partial H}{\partial p}(\phi(t), \nabla_x V(\phi(t))), \\ H(\phi(t), \nabla_x V(\phi(t))) = 0. \end{cases}$$

Assume that for all $i = 1, \ldots, n$, $\phi_i(t) = \frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}}$. Then, by Lemma 3, we have: $\min_{p_i \in \mathbb{R}} H_i(\phi(t), p_i) = 0$ for all i.

Thereby, $H(\phi(t), \nabla_x V(\phi(t))) = 0$ if and only if for all i $H_i(\phi(t), \partial_{x_i} V(\phi(t)) = \min_{p_i \in \mathbb{R}} H_i(\phi(t), p_i) = 0$, which implies: $\frac{\partial H}{\partial p}(\phi(t), \nabla_x V(\phi(t))) = \dot{\phi}(t) = 0$. The lemma is proved. □

Proof of Theorem 4.(ii) From the Corollary 1, if $\nabla_x V(\phi(t)) = 0$, the trajectory is a relaxation trajectory, along which the gradient is uniformly equal to zero. The condition (C) implies that it is reduced to a single point: $\dot{\phi}(t) = 0$. Conversely, with the same reasoning that for the proof of (i):

$$\dot{\phi}(t) = 0 \implies L(\phi(t), \dot{\phi}(t)) = 0 \implies \forall i : \dot{\phi}_i(t) = d_i \left(\frac{k_{on,i}(\phi(t))}{k_{on,i}(\phi(t)) + k_{off,i}} - \phi_i(t) \right).$$

We recognize the equation of a relaxation trajectory, which implies: $\nabla_x V(\phi(t)) = 0$. Thus, for any optimal trajectory satisfying the system (16) associated to a solution V of the Eq. (25) satisfying the condition (C), we have for any time t:

$$\dot{\phi}(t) = 0 \iff \nabla_x V(\phi(t)) = 0.$$

From the equivalence proved in (i), the condition (C) then implies that the gradient of V vanishes only on the stationary points of the system (4). □

Proof of Theorem 4.(iii) As a consequence of (ii), for any optimal trajectory ϕ_t associated to a solution V of (25) which satisfies the condition (C), we have for all $t > 0$:

$$\dot{\phi}(t) = 0 \iff \nabla_x V(\phi(t)) = 0.$$

Then, if there exists $t > 0$ such that $\dot{\phi}(t) \neq 0$, it cannot be equal to the drift of a relaxation trajectory, defined by the deterministic system (4), which is known to be the unique velocity field for which the Lagrangian vanishes (from Theorem (3)). Then it implies:

$$\dot{\phi}(t) \neq 0 \implies L(\phi(t), \dot{\phi}(t)) \neq 0.$$

The relation (43), combined to the fact that the Lagrangian is always nonnegative allow to conclude:

$$\dot{\phi}(t) \neq 0 \implies \sum_{i=1}^{n} \partial_{x_i} V(\phi(t))\dot{\phi}_i(t) = \frac{\partial V}{\partial t}(\phi(t)) > 0.$$

Thus, the function V strictly increases on these trajectories.

Furthermore, on any relaxation trajectory $\phi_r(t)$, from the inequality (15) we have for any times $T1 < T2$:

$$0 = \int_{T_1}^{T_2} L(\phi_f(t), \dot{\phi}_r(t))dt \geq V(\phi_r(T_2)) - V(\phi_r(T_1)).$$

The equality holds between $T1$ and $T2$ if and only if for any $t \in [T1, T2]$: $L(\phi(t), \dot{\phi}(t)) = 0$. In that case, from Theorem 3the drift of the trajectory is necessarily the drift of a relaxation trajectory between the two timepoints and then, from Corollary 1, $\nabla_x V = 0$ on the set of points $\{\phi(t), t\mathbb{R}^+\}$, which is excluded by the condition (C) (when this set is not reduced to a single point). Thus, if $\phi(T_1) \neq \phi(T_2)$, we have: $V(\phi(T_1)) > V(\phi(T_2))$.

By definition, for any basin Z_i and for all $x \in Z_i$ there exists a relaxation trajectory connecting x to the associated attractor X_{eq,Z_i}. So $\forall x \in Z_i,\ V(x) > V(X_{eq,Z_i})$. □

Proof of Theorem 4.(iv) Let V be a solution of (25) satisfying the condition (C). We consider trajectories solutions of the system defined by the drift $\dot{\phi}(t) = -\nabla_p H(\phi(t), \nabla_x V(\phi(t)))$. We recall that from (iii), the condition (C) ensures that

V decreases on these trajectories, and that for all $t_1 < t_2$: $V(\phi(t_1)) - V(\phi(t_2)) = Q(\phi(t_2), \phi(t_1)) > 0$. Then, the hypothesis $\lim_{x \to \partial\Omega} V(x) = \infty$ ensures that such trajectories cannot reach the boundary $\partial\Omega$: if it was the case, we would have a singularity inside Ω, which is excluded by the condition $V \in C^1(\Omega, \mathbb{R})$. The same reasoning also ensures that there is no limit cycle or more complicated orbits for this system.

Recalling that from (i) and (ii), the fixed point of this system are reduced to the points where $\nabla_x V = 0$ on Ω, we conclude that for all $x \in \Omega$, there exists a fixed point $a \in \Omega$, satisfying $\nabla_x V(a) = 0$, such that a trajectory solution of this system converges to a, i.e: $V(x) - V(a) = Q(a, x)$.

As from the inequality (15), we have for every point a the relation $V(x) - V(a) \leq Q(a, x)$, the previous equality corresponds to a minimum and we obtain the formula:

$$\forall x \in \Omega,\ V(x) = \min_{\{a \in \Omega | \nabla_x V(a) = 0\}} V(a) + Q(a, x).$$

□

Proof of Theorem 4.(v) Let V be a solution of (25) satisfying the condition (C). We denote by ν_V the drift of the optimal trajectories ϕ_t on $[0, T]$ satisfying the system (16) associated to V: $\forall t \in [0, T]$, $\dot{\phi}(t) = \nabla_p H(\phi(t), \nabla_x V(\phi(t))) = \nu_V(\phi(t))$. We call trajectories solution of this system *reverse fluctuations* trajectories.

For any basin Z_i associated to the stable equilibrium of the deterministic system X_{eq,Z_i}, we have:

- From (i), $\nu_V(X_{eq,Z_i}) = 0$ and $\forall x \in Z_i \setminus X_{eq,Z_i} : \nu_V(x) \neq 0$.
- From (iii), we know that V increases on these trajectories: $\forall x \in Z_i \setminus X_{eq,Z_i} : \langle \nabla_x V(x), \nu_V(x) \rangle > 0$.
- From (iii), we also have: $V(X_{eq,Z_i}) = \min_{x \in Z_i} V(x)$.

Without loss of generality (since we only use $\nabla_x V$), we can assume $V(X_{eq,Z_i}) = 0$. We have then: $\forall x \in Z_i \setminus \{X_{eq,Z_i}\}$, $V(x) > 0$. Moreover, since we have assumed that X_{eq,Z_i} is isolated, there exists $\delta_V > 0$ such that Z_i contains a ball $B(X_{eq,Z_i}, \delta_V)$. Therefore, V reaches a local minimum at X_{eq,Z_i}. Conversely if V reaches a local minimum at a point $\bar{x}$, then $\bar{x}$ is necessarily an equilibrium (from (ii)), and the fact that V strictly decreases on the relaxation trajectories ensures that it is a Lyapunov function for the deterministic system, and then that $\bar{x}$ is a stable equilibrium. The stable equilibria of the deterministic system are thereby exactly the local minima of V, and for any attractor X_{eq,Z_i}, V is also a Lyapunov function for the system defined by the drift $-\nu_V$, for which X_{eq,Z_i} is then a locally asymptotically stable equilibrium. Thereby, stable equilibria of the deterministic system are also stable equilibria of the system defined by the drift $-\nu_V$.

It remains to prove that no unstable equilibria the deterministic system is stable for the system defined by the drift $-\nu_V$. Let $\bar{x}$ be an unstable equilibrium of the relaxation system, then V does not reach a local minimum at $\bar{x}$. Therefore, as close as we want of $\bar{x}$ there exists x such that $V(x) < V(\bar{x})$. We recall that reverse fluctuations trajectories ϕ_t starting from such a point and remaining in Ω will have $V(\phi(t))$ stricly decreasing: by Lyapunov La Salle principle, they shall be attracted towards the set

$\{y, \nabla\langle V(y), \nu_V(y)\rangle = 0\}$, which contains from (iii) only the critical points of V, which are from (i) the equilibria of both deterministic and reverse fluctuation systems. In particular, either ϕ_t leaves Ω (and the equilibrium is unstable) or ϕ_t converges to another equilibrium (since they are isolated) and this contradicts the stability. So we have proved that stable equilibria of both systems are the same.

We then obtain that for any Z_i, there exists δ_V such that:

$$\forall x \in B(X_{eq,Z_i}, \delta_V),\ \exists \phi_t \in C^1(\Omega) : \{\phi(0) = x,\ \forall t \in [0, T] : \dot{\phi}(t) = -\nu_V(\phi(t)),\ , \phi(T) \underset{T\to\infty}{\to} X_{eq,Z_i}\}.$$

Reverting time, any point of $B(X_{eq,Z_i}, \delta_V)$ can then be reached from any small neighborhood of X_{eq,Z_i}. We deduce from Lemma 1 that:

$$\forall x \in B(X_{eq,Z_i}, \delta_V) :\ Q(X_{eq,Z_i}, x) = V(x) - V(X_{eq,Z_i}).$$

Applying exactly the same reasoning to another function $\tilde{V}$ solution of (25) and satisfying (C), this ensures that $V(x) - \tilde{V}(x) = V(X_{eq,Z_i}) - \tilde{V}(X_{eq,Z_i})$, at least for $x \in B(X_{eq,Z_i}, \min(\delta_V, \delta_{\tilde{V}}))$.

We recall that that from Lemma 2, two optimal trajectories $\phi_t, \tilde{\phi}_t$ solutions of the system (16), associated respectively to two solutions V and $\tilde{V}$ of the Eq. (25), cannot cross each other without satisfying $\nabla_x V = \nabla_x \tilde{V}$ along the whole trajectories. Thereby, we can extend the equality $\nabla_x V = \nabla_x \tilde{V}$ on the basins of attraction associated to the stable equilibrium X_{eq,Z_i} for both systems defined by the drifts $-\nu_V$ or $-\nu_{\tilde{V}}$. Thus, we have proved that the basins associated to the attractors are the same for both systems. We denote $(Z_i^f)_{Z_i \in Z}$ these common basins.

Under the assumption 2. of the theorem, we obtain by continuity of V that for every pair of basin (Z_i^f, Z_j^f), $V(X_{eq,Z_i}) - \tilde{V}(X_{eq,Z_i}) = V(X_{eq,Z_j}) - \tilde{V}(X_{eq,Z_j})$. It follows that under this assumption, there exists a constant $c \in \mathbb{R}$ such that for every attractor X_{eq,Z_i}:

$$V(X_{eq,Z_i}) = \tilde{V}(X_{eq,Z_i}) + c. \tag{44}$$

Moreover, the assumption 1. ensures that from Theorem 4.(iv), there exists a fixed point $a_1 \in \Omega$ (with $\nabla_x V(a_1) = 0$), such that a trajectory solution of the system defined by the drift $-\nu_V$ converges to a_1, i.e: $V(x) - V(a_1) = Q(a_1, x)$. On one side, if a_1 is unstable, it necessarily exists on any neighborhood of a_1 a point x_2 such that $V(x_2) < V(a_1)$. As for all $y \in \Omega$, $Q(\cdot, y)$ is positive definite, we have then another fixed point $a_2 \neq a_1$ such that $V(x_2) = Q(a_2, x_2) + V(a_2)$. We obtain: $V(x) > h(x, a_1) + Q(a_2, x_2) + V(a_2)$. On the other side, by continuity of the function $Q(a_2, \cdot)$, for every $\delta_1 > 0$, x_2 can be chosen close enough to a_1 such that: $Q(a_2, x_2) \geq Q(a_2, a_1) - \delta_1$. We obtain:

$$\forall \delta_1 > 0,\ \exists x_2,\ \exists a_2 \neq a_1 :\ V(x) > Q(a_1, x) + Q(a_2, a_1) + V(a_2) - \delta_1.$$

Repeating this procedure until reaching a stable equilibrium at a step N, which is necessarily finite because we have by assumption a finite number of fixed points, we

obtain the inequality

$$\forall x \in \Omega,\ \forall \delta > 0,\ \exists (a_k)_{n=1,\dots,N} :\ V(x) > Q(a_1, x) + V(a_N) + \sum_{k=1}^{N-1} Q(a_{k+1}, a_k) - \delta,$$

where every a_k denotes a fixed point and a_N is an attractor. Using the triangular inequality satisfied by Q, and passing to the limit $\delta \to 0$, we find that $V(x) - V(a_N) \geq Q(a_N, x)$. Moreover, from the inequality (15), we have necessarily $\tilde{V}(x) - \tilde{V}(a_N) \leq Q(a_N, x)$. It then follows from (44) that $\tilde{V}(x) + c \leq V(x)$.

Applying exactly the same reasoning for building a serie of fixed point $(\tilde{a}_k)_{n=1,\dots,\tilde{N}}$ such that $\tilde{a}_{\tilde{N}}$ is an attractor and $\tilde{V}(x) - \tilde{V}(\tilde{a}_{\tilde{N}}) \geq Q(\tilde{a}_{\tilde{N}}, x)$, we obtain $\tilde{V}(x) \geq V(x) - c$. We can conclude:

$$\forall x \in \Omega :\ V(x) = \tilde{V}(x) + c.$$

□

Proof of Theorem 5 First, we prove the following lemma:

Lemma 4 *$\forall i \in \{1, \dots, n\}$, we have:*

(i) *$\exists \delta_l > 0,\ \exists \eta_l > 0$, such that $\forall x, y \in \Omega$, if $y_i < x_i \leq \delta_l$, then we have:*

$$Q(x, y) \geq \eta_l \ln \frac{x_i}{y_i}.$$

(ii) *$\exists \delta_r < 1,\ \exists \eta_r > 0$, such that $\forall x, y \in \Omega$, if $y_i > x_i \geq \delta_r$, then we have:*

$$Q(x, y) \geq \eta_r \ln \frac{1 - x_i}{1 - y_i}.$$

Proof (i) We denote $m_i = \min_{x \in \Omega} k_{on,i}(x)$. We have $m_i > 0$ by assumption. We choose a real number δ which satisfies these two conditions:

1. $0 < \delta_l < \frac{m_i}{d_i(m_i + k_{off,i})}$,
2. $\sqrt{k_{off,i}\delta_l} - \sqrt{m_i(1 - \delta_l)} \leq -\sqrt{\frac{m_i}{2}}$.

On the one hand, we recall that the function $v_i \to L_i(x, v_i) = \left(\sqrt{k_{off,i}\frac{v_i + d_i x_i}{d_i}} - \sqrt{k_{on,i}(x)\frac{d_i(1-x_i) - v_i}{d_i}}\right)$ is convex and vanishes only on $v_i = d_i\left(\frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} - x_i\right)$. Then, $L_i(x, \cdot)$ is decreasing on $[-d_i x_i, d_i\left(\frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} - x_i\right)]$.

On the other hand, for all $x \in \Omega$, if $x_i \leq \delta_l$, we have necessarily: $d_i\left(\frac{k_{on,i}(x)}{k_{on,i}(x) + k_{off,i}} - x_i\right) \geq d_i(m_i - \delta_l) > 0$ from the condition 1. Then, we obtain that for all $x \in \Omega$, if $x_i \leq \delta_l$:

$$\forall v_i \in [-d_i x_i, 0] :\ L_i(x, v_i) \geq L_i(x, 0).$$

From the condition 2., we also see that for all $x \in \Omega$, if $x_i \leq \delta_l$:

$$\sqrt{k_{off,i} x_i} - \sqrt{k_{on,i}(x)(1 - x_i)} \leq \sqrt{k_{off,i} \delta_l} - \sqrt{m_i (1 - \delta_l)} \leq -\sqrt{\frac{m_i}{2}},$$

which implies:

$$L_i(x, 0) = \left(\sqrt{k_{off,i} x_i} - \sqrt{k_{on,i}(x)(1 - x_i)}\right)^2 \geq \frac{m_i}{2}.$$

Then, we obtain that for any admissible trajectory $\phi_t \in C_{0T}^{1,pw}(\Omega)$ (i.e with a velocity in $\Omega^v(\phi(t))$ at all time) and such that $\phi(0) = x$ and $\phi(T) = y$, if $y_i < x_i \leq \delta_l$ we have:

$$\begin{aligned} J_T(\phi) &= \int_0^T L(\phi(t), \dot{\phi}(t))dt \geq \int_0^T L_i(\phi(t), \dot{\phi}_i(t))dt \geq \int_0^T \mathbb{1}_{\{\dot{\phi}_i(t) \leq 0,\, \phi_i(t) \leq \delta_l\}} L_i(\phi(t), \dot{\phi}_i(t))dt, \\ &\geq \int_0^T \mathbb{1}_{\{\dot{\phi}_i(t) \leq 0,\, \phi_i(t) \leq \delta_l\}} L_i(\phi(t), 0)dt \geq \int_0^T \mathbb{1}_{\{\dot{\phi}_i(t) \leq 0,\, \phi_i(t) \leq \delta_l\}} \frac{m_i}{2} dt, \\ &\geq \frac{m_i}{2} \sum_{k=1}^{l} (t_{r,k} - t_{l,k}), \end{aligned} \tag{45}$$

where we denote $\{[t_{l,k}, t_{r,k}]\}_{k=1,\ldots,l}$ the l intervals on which the velocity $\dot{\phi}_i(t) < 0$ and $\phi_i(t) < \delta_l$ on the interval $[0, T]$. As we now by assumption that $\phi_i(0) = x_i \leq \delta_l$ and $\phi_i(T) = y_i < \phi_i(0)$, this set of intervals cannot be empty.

Moreover, for every $k = 1, \ldots, l$, we have by assumption: $\forall t \in [t_{l,k}, t_{r,k}]$, $-d_i \phi_i(t) \leq \dot{\phi}_i(t) \leq 0$. Then:

$$\phi_i(t_{r,k}) \geq \phi_i(t_{l,k}) e^{-d_i (t_{d,k} - t_{g,k})}.$$

As by definition, for every $k = 1, \ldots, l-1$, $\dot{\phi}_i(t) \geq 0$ on $[t_{r,k}, t_{l,k+1}]$, we have $\phi_i(t_{l,k+1}) \geq \phi_i(t_{r,k})$ (because $\dot{\phi}_i(t) \geq 0$ on $[t_{r,k}, t_{l,k+1}]$). Finally, we obtain:

$$\phi(T) = y_i \geq e^{-d_i \left(\sum_{k=1}^{l} t_{r,k} - t_{l,k} \right)} \phi_i(0) = x_i,$$

which implies:

$$\sum_{k=1}^{l} (t_{r,k} - t_{l,k}) \geq \frac{1}{d_i} \ln \frac{x_i}{y_i}.$$

This last inequality combined with (45) allows to conclude:

$$J_T(\phi) \geq \frac{m_i}{2d_i} \ln \frac{x_i}{y_i}.$$

Thus, if δ_l satisfies conditions 1. and 2., and fixing $\eta_l = \frac{m_i}{2d_i} > 0$, for every $x, y \in \Omega$ such that $y_i < x_i \leq \delta_l$ we have:

$$Q(x, y) \geq \eta_l \ln \frac{x_i}{y_i}.$$

(ii) Denoting $M_i = \max_{x \in \Omega} k_{on,i}(x)$ (which exists by assumption), we chose this time the real number δ_r in order to satisfy these two conditions:

1. $1 > \delta_r > \frac{M_i}{d_i(M_i + k_{off,i})}$,
2. $\sqrt{k_{off,i}\delta_r} - \sqrt{M_i(1-\delta_r)} \geq \sqrt{\frac{k_{off,i}}{2}}$,

and we fix $\eta_r = \frac{k_{off,i}}{2d_i}$. The rest consists in applying exactly the same reasoning than for the proof of (i) in a neighborhood of 1 instead of 0. □

We deduce immediately the following corollary:

Corollary 2 $\forall x \in \Omega, \lim_{y \to \partial\Omega} Q(x, y) = \infty.$

Let us denote $V \in C^1(\Omega, \mathbb{R})$ a solution of the Eq. (25), which satisfies the condition (C). From the proof of Theorem 4.(v), we know that for any attractor X_{eq,Z_i}, there exists a ball $B(X_{eq,Z_i}, \delta) \subset Z_i^f$, where Z_i^f is the basin of attraction of X_{eq,Z_i} for the system defined by the drift $-\nu_V(\cdot) = -\frac{\partial}{\partial p} H(\cdot, \partial_x V(\cdot))$. Moreover, as V decreases on trajectories solutions of this system, the set $Z_i^V = \{x \in Z_i \mid V(x) \leq \min_{y \in \partial Z_i} V(y)\}$ is necessarily stable: we have $Z_i^V \subset Z_i^f$.

We deduce that:

$$\forall x \in Z_i^V, \exists \phi_t \in C^1(\Omega) : \{\phi(0) = x, \forall t \in [0, T] : \dot{\phi}(t) = -\nu_V(\phi(t)), \phi(T) \underset{T \to \infty}{\to} X_{eq,Z_i}\},$$

and in that case $Q(X_{eq,Z_i}, x) = V(x) - V(X_{eq,Z_i})$. If there existed $y \in \partial\Omega \cap \overline{Z}_i^V$, we would have, by continuity of V and from Corollary 2:

$$\lim_{x \to y} (V(x) - V(X_{eq,Z_i}) = \lim_{x \to y} Q(X_{eq,Z_i}, x) = \infty.$$

It would imply that $\min_{y \in \partial Z_i} V(y) = \infty$, which is impossible when $\partial Z_i \neq \partial\Omega$, which is necessarily the case when there is more than one attractor.

Thus, there exists at least one point x^i on the boundary $\partial Z_i \setminus \partial\Omega$, such that for any neighborhood of X_{eq,Z_i}, there exists a fluctuation trajectory starting inside and converging to x^i.

We recall that we assume that $\Omega \subset \bigcup_{Z_i \in Z} \overline{Z}_i$. Then we have $\partial Z_i \setminus \partial\Omega = \bigcup_{k \neq i} \{\partial Z_i \cap \partial Z_k\}$ and there exists Z_j such that: $x^i \in \partial Z_i \cap \partial Z_j = \partial Z_i \setminus R_{ij}$. We obtain, by continuity of V:

$$Q_{R_{ij}}(X_{eq,Z_i}, \partial Z_i \cap \partial Z_j) = \min_{y \in \partial Z_i \cap \partial Z_j} V(y) - V(X_{eq,Z_i}) = V(x^i) - V(X_{eq,Z_i}).$$

It remains to prove that under the assumption (A) of the theorem, $x^i = x^{ij}_{un}$. On one hand, from Theorem 4.(iii), V decreases on the relaxation trajectories. On the other hand, if every relaxation trajectories starting in $\partial Z_i \cap \partial Z_j$ stay inside, they necessarily converge from any point of $\partial Z_i \cap \partial Z_j$ to a saddle point (also in $\partial Z_i \cap \partial Z_j$). Then, the minimum of V in $\partial Z_i \cap \partial Z_j$ is reached on the minimum of V on X^{ij}_{un} (the set of all the saddle points in $\partial Z_i \cap \partial Z_j$), which is x^{ij}_{un}. Thus, if every relaxation trajectories starting in $\partial Z_i \cap \partial Z_j$ stay inside, then $X_{Z_i} \in \partial Z_j$ implies $X_{Z_i} = x^{ij}_{un}$. The theorem is proved. □

I Algorithm to find the saddle points

We develop a simple algorithm using the Lagrangian associated to the fluctuation trajectories (28) to find the saddle points of the deterministic system (4). This Lagrangian is a nonnegative function which vanishes only at the equilibria of this system. Then, if there exists a saddle point connecting two attractors, this function will vanish at this point. Starting on a small neighborhood of the first attractor, we follow the direction of the second one until reaching a maximum on the line (see 15a). Then, we follow different lines, in the direction of each other attractor for which the Lagrangian function decreases (at least, the direction of the second attractor (see 15b)), until reaching a local minimum. We then apply a gradient descent to find a local minimum (see 15c). If this minimum is equal to 0, this is a saddle point, if not we repeat the algorithm from this local minimum until reaching a saddle point or an attractor. Repeating this operation for any ordered couple of attractors $(X_{eq,Z_i}, (X_{eq,Z_j})_{Z_i,Z_j \in Z, i \neq j}$, we are likely to find most of the saddle points of the system. This method is described in pseudo-code in Algorithm 1.

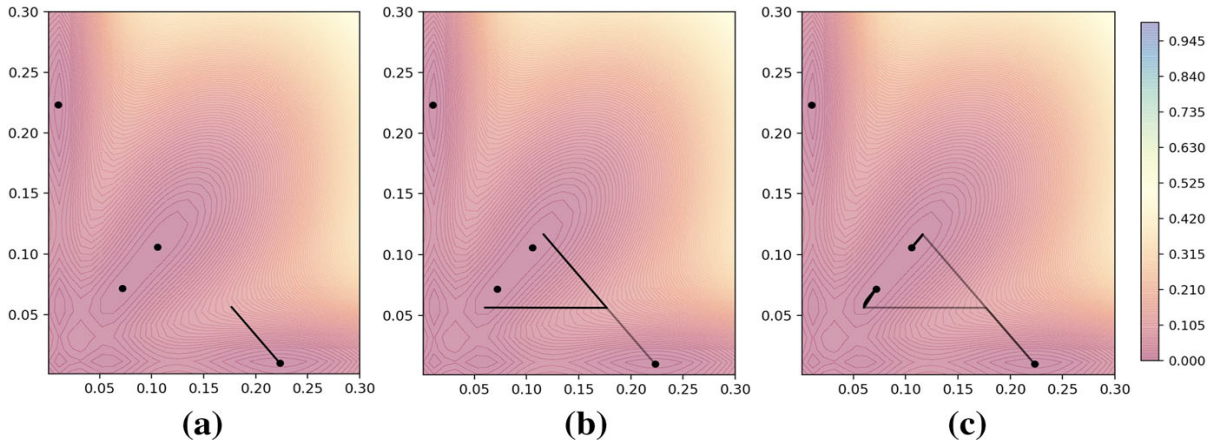


Fig. 15 Saddle-point algorithm between two attractors. The color map corresponds to the Lagrangian function associated to the fluctuation trajectories

Algorithm 1 Find the list of saddle points: list-saddle-points

Require: • The list of attractors: list-attractors $= (X_{eq,Z_i})_{Z_i \in Z}$
- The Lagrangian function to minimize on the saddle points: $Lag : \mathbb{R}^n \to \mathbb{R}^+$
- A gradient descent function, finding a local minimum of Lag from a point x: gradient-descent(x)
- A subdivision coefficient: $\alpha \ll 1$

while $X_{eq,Z_i} \in$ list-attractors **do**
  $X = X_{eq,Z_i}$
  while $X_{eq,Z_j} \in$ list-attractors$\backslash X_{eq,Z_i}$ **do**
    $Lag0 = Lag(X)$
    $X \leftarrow X + \alpha(X_{eq,Z_i} - X_{eq,Z_j})$
    while $Lag(X) >= Lag0$ **do**
      $Lag0 \leftarrow Lag(X)$
      $X \leftarrow X + \alpha(X_{eq,Z_i} - X_{eq,Z_j})$
    end while
    while $X_{eq,Z_j} \in$ list-attractors$\backslash X_{eq,Z_i}$ **do**
      while $Lag(X) < Lag0$ **do**
        $Lag0 \leftarrow Lag(x)$
        $X \leftarrow X + \alpha(X_{eq,Z_i} - X_{eq,Z_j})$
      end while
      $X_0 =$ gradient-descent(X)
      if $Lag(X_0) = 0$ **then**
        list-saddle-points $\leftarrow X_0$
      end if
    end while
  end while
end while

J Applicability of the method for non-symmetric networks and more than two genes

We present in Figs. 16b and 17b an analogy of Fig. 8, which was presented for the toggle-switch network, for two non-symmetric networks of respectively 3 and 4 genes. The networks are presented on the left-hand side of Figs. 16b and 17b: the red arrows represent the inhibitions and the green arrows represent the activations between genes. A typical random trajectory for each network is presented in Figs. 16a and 17a.

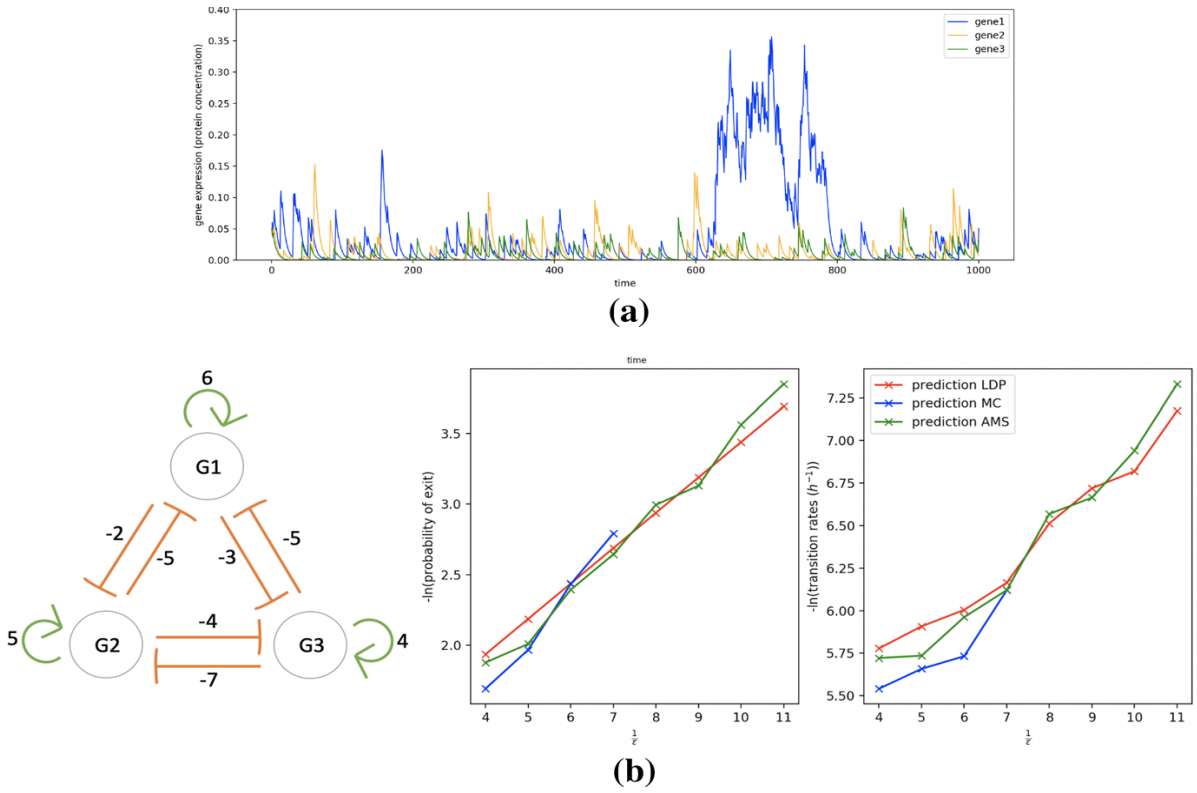


Fig. 16 **a** A random trajectory associated to the non-symmetric toggle-switch network of 3 genes with $\varepsilon = 1/8$. The network is associated to 2 attractors only: Z_{---} all the genes inactive, and Z_{+--} gene 1 active and gene 2 and gene 3 inactive due to the inhibitions. **b** Analogy of Fig. 8 between Z_{---} and Z_{+--}. We see that the analytical approximations of the transition rates are very accurate although we had no theoretical evidence for the trajectories computed by the method presented in Sect. 5.3 to be optimal

We recall that we build the LDP approximation (in red) by using the cost of the trajectories satisfying the system (28) between the attractors and the saddle points of the system (4). The cost of these trajectories is known to be optimal when there exists a solution V of the Eq. (25) which verifies the relations (26), which can generally happen only under symmetry conditions. This is not the case nor for the 3 genes network of Fig. 16b when there is no symmetry between the interactions, neither for the 4 genes network of Fig. 17b. Then, we could expect that these LDP approximations would be far from the Monte-Carlo and AMS computations, especially for the 4 genes network, since we have no symmetry between the interactions, not only in value but also in sign. However, we observe that the approximations given by our method seem to remain relatively accurate.

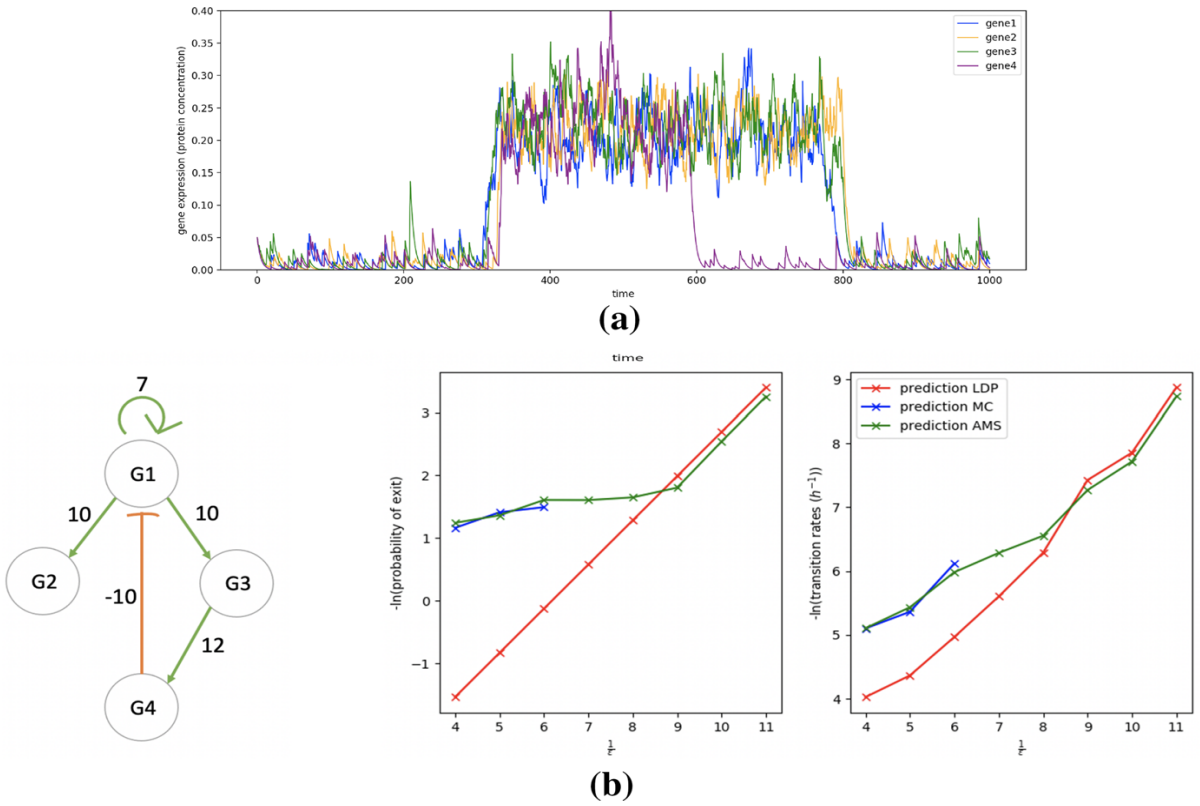


Fig. 17 **a** A random trajectory associated to the non-symmetric 4 genes network with $\varepsilon = 1/8$. The network is associated to 3 attractors: Z_{----} all the genes inactive, Z_{+++-} genes 1–2–3 active and gene 4 inactive, and Z_{++++} all the genes active. **b**: Analogy of Fig. 8 between Z_{++++} and Z_{+++-}. The analytical approximations seem to become accurate from $\varepsilon \simeq 1/9$

References

Albayrak C et al (2016) Digital quantification of proteins and mRNA in single mammalian cells. Mol Cell 61(6):914–924

Antolovic V et al (2017) Generation of single-cell transcript variability by repression. Curr Biol 27(12):1811-1817 e3

Berglund N (2011) "Kramers' law: validity, derivations and generalisations". arXiv:1106.5799

Bizzarri M, Masiello MG, Giuliani A, Cucina A (2018) Gravity constraints drive biological systems toward specific organization patterns: commitment of cell specification is constrained by physical cues. Bioessays 40(1):1700138

Bouchet F, Reygner J (2016) Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes. Annales Henri Poincaré. 17(12):3499–3532

Bouchet F et al (2016) Large deviations in fast-slow systems. J Stat Phys 162(4):793–812

Brackston RD, Wynn A, Stumpf MP (2018) Construction of quasipotentials for stochastic dynamical systems: an optimization approach. Phys Rev E 98(2):022136

Braun E (2015) The unforeseen challenge: from genotype-to-phenotype in cell populations. Rep Prog Phys 78(3):036602

Bréhier C-E, Lelièvre T (2019) On a new class of score functions to estimate tail probabilities of some stochastic processes with adaptive multilevel splitting. Chaos Interdiscip J Nonlinear Sci 29(3):033126

Bréhier C-E et al (2016) Unbiasedness of some generalized adaptive multilevel splitting algorithms. Ann Appl Probab 26(6):3559–3601

Bressloff PC (2014) Stochastic processes in cell biology, vol 41. Springer, Berlin

Bressloff PC, Faugeras O (2017) On the Hamiltonian structure of large deviations in stochastic hybrid systems. J Stat Mech Theory Exp 2017(3):033206

Budhiraja A, Dupuis P (2019) Multilevel splitting analysis and approximation of rare events. Springer, Berlin, pp 439–469
Cérou F et al (2011) A multiple replica approach to simulate reactive trajectories. J Chem Phys 134(5):054108
Chu BK et al (2017) Markov state models of gene regulatory networks. BMC Syst Biol 11(1):14
Clevers H et al (2017) What is your conceptual definition of "cell type" in the context of a mature organism? Cell Syst 4:255–259
Cohen JE (1981) Convexity of the dominant eigenvalue of an essentially nonnegative matrix. Proc Am Math Soc 81(4):657–658
Coskun AF, Eser U, Islam S (2016) Cellular identity at the single-cell level. Mol BioSyst 12(10):2965–2979
Dean T, Dupuis P (2009) Splitting for rare event simulation: a large deviation approach to design and analysis. Stoch Process Appl 119(2):562–587
Dembod A, Zeltouni O, Fleischmann K (1996) Large deviations techniques and applications. Jahresber Dtsch Mathematiker Ver 98(3):18–18
Evans LC (2010) Partial differential equations. American Mathematical Society, Providence. ISBN: 9780821849743 0821849743
Faggionato A, Gabrielli D, Crivellari MR (2009) Non-equilibrium thermodynamics of piecewise deterministic Markov processes. J Stat Phys 137(2):259
Fathi A (2008) Weak KAM theorem in Lagrangian dynamics preliminary version number 10. by CUP
Freidlin MI, Wentzell AD (2012) Random perturbations of dynamical systems. Third. Vol. 260. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Heidelberg
Gao NP et al (2020) Universality of cell differentiation trajectories revealed by a reconstruction of transcriptional uncertainty landscapes from single-cell transcriptomic data. bioRxiv
Gobet E (2000) Weak approximation of killed diffusion using Euler schemes. Stoch Process Appl 87(2):167–197 (**issn: 0304-4149**)
Guillemin A et al (2019) Drugs modulating stochastic gene expression affect the erythroid differentiation process. PLoS ONE 14(11):e0225166
Gupta PB et al (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146(4):633–44
Herbach U et al (2017) Inferring gene regulatory networks from single-cell data: a mechanistic approach. BMC Syst Biol 11(1):105 (**issn: 1752-0509**)
Heymann M, Vanden-Eijnden E (2008) The geometric minimum action method: A least action principle on the space of curves. Commun Pure Appl Math J Issued Courant Inst Math Sci 61(8):1052–1117
Huang S, Ingber DE (2007) A non-genetic basis for cancer progression and metastasis: self-organizing attractors in cell regulatory networks. Breast Dis 26(1):27–54
Huang S et al (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys Rev Lett 94(12):128701
Kauffman S (2004) A proposal for using the ensemble approach to understand genetic regulatory networks. J Theor Biol 230(4):581–590
Kifer Y (2009) Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging. American Mathematical Soceity, New York
Ko MSH (1991) A stochastic model for gene induction. J Theor Biol 153(2):181–194
Kurtz TG, Swanson J (2019) Finite Markov chains coupled to general Markov processes and an application to metastability. arXiv:1906.03212
Li G-W, Xie XS (2011) Central dogma at the single-molecule level in living cells. Nature 475(7356):308–315
Li Y, Duan J, Liu X (2021) Machine learning framework for computing the most probable paths of stochastic dynamical systems. Phys Rev E 103(1):012124
Lin YT, Galla T (2016) Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models. J R Soc Interface 13(114):20150772
Lv C et al (2014) Constructing the energy landscape for genetic switching system driven by intrinsic noise. PLoS ONE 9(2):e88167
Ma Z, Leijon A (2009) Beta mixture models and the application to image classification. In: 2009 16th IEEE international conference on image processing (ICIP). IEEE, pp 2045–2048

Mar JC (2019) The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond. Biophys Rev 11:89–94 (**Please check and confirm the inserted volume number is correct for the reference Mar (2019).**)
Mohammed H et al (2017) Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. Cell Rep 20(5):1215–1228
Mojtahedi M et al (2016) Cell fate decision as high-dimensional critical state transition. PLoS Biol 14(12):e2000640
Moon KR et al (2018) Manifold learning-based methods for analyzing single-cell RNA-sequencing data. Curr Opin Syst Biol 7:36–46
Moris N, Arias AM (2017) The hidden memory of differentiating cells. Cell Syst 5(3):163–164
Moris N, Pina C, Arias AM (2016) Transition states and cell fate decisions in epigenetic landscapes. Nat Rev Genet 17(11):693–703
Morris SA (2019) The evolving concept of cell identity in the single cell era. Development 146(12):1–5
Moussy A et al (2017) Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. PLoS Biol 15(7):e2001867
Newby JM, Keener JP (2011) An asymptotic analysis of the spatially inhomogeneous velocity-jump process. Multiscale Model Simul 9(2):735–765
Pakdaman K, Thieullen M, Wainrib G (2012) Asymptotic expansion and central limit theorem for multiscale piecewise-deterministic Markov processes. Stoch Process Appl 122(6):2292–2318
Papanicolaou GC (1975) Asymptotic analysis of transport processes. Bull Am Math Soc 81(2):330–392
Pearce P et al (2019) Learning dynamical information from static protein and sequencing data. Nat Commun 10(1):1–8
Peccoud J, Ycart B (1995) Markovian modeling of gene-product synthesis. Theor Popul Biol 48(2):222–234
Pratapa A et al (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods 17(2):147–154
Richard A et al (2016) Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLoS Biol 14(12):e1002585
Richard A et al (2019) Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements. PLoS ONE 14(9):e0221472
Schwanhäusser B et al (2011) Global quantification of mammalian gene expression control. Nature 473(7347):337
Semrau S et al (2017) Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat Commun 8(1):1–16
Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Natl Acad Sci 105(45):17256–17261
Stumpf MP et al (2017) Stem cell differentiation as a non-Markov stochastic process. Cell Syst 5:268–282
Suter DM et al (2011) Mammalian genes are transcribed with widely different bursting kinetics. Science 332(6028):472–474
Tong M et al (2018) Transcriptomic but not genomic variability confers phenotype of breast cancer stem cells. Cancer Commun (Lond) 38(1):56
Waddington CH (1957) The strategy of the genes. Routledge, London
Wang J et al (2010) The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. Biophys J 99(1):29–39
Wang J et al (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. Proc Natl Acad Sci 108(20):8257–8262
Wheat JC, Sella Y, Willcockson M, Skoultchi AI, Bergman A, Singer RH, Steidl U (2020) Single-molecule imaging of transcription dynamics in somatic stem cells. Nature 583(7816):431–436
Zhou JX et al (2012) Quasi-potential landscape in complex multi-stable systems. J R Soc Interface 9(77):3539–3553
Zhou JX et al (2014) Nonequilibrium population dynamics of phenotype conversion of cancer cells. PLoS ONE 9(12):e110714
Zhou P, Li T (2016) Construction of the landscape for multi-stable systems: potential landscape, quasi-potential, A-type integral and beyond. J Chem Phys 144(9):094109

Appendix K - Stationary distribution of the phenomenological model

In this new appendix, we prove that the stationary distribution characterizing the phenomenological model, described in Figure 11 (but slightly modifying the Markov chain on the basins), appears as the mixture of Beta distributions stated in (33).
We consider $\varepsilon = 1$ for simplifying the notations. We consider a PDMP system on the $2n+1$ variables $(X_i, E_i)_{i=1,\cdots,n}$ and Z. E_i and Z are then discrete variables modeling the promoter state pf gene i and the cellular type and X_i a continuous variable modeling the protein state. We consider that Z follows a Markov chain on the discrete space $(Z_1, \cdots, Z_{N_Z})$, where N_Z denotes the number of basins, with exponential rates defined by:

$$\forall (k,l) \in \{1, \cdots, N_Z\}^2 : a_{Z_k,Z_l}(x) = \frac{\tilde{a}_{Z_k,Z_l}}{\prod_i Beta\left(\frac{k_{on,i}(X_{eq,Z_k})}{d_i}, \frac{k_{off,i}}{d_i}\right)(x)}. \tag{2.1}$$

This process is well defined, by the same arguments as the ones used for justifying the process (1.15). We denote μ_Z the stationary distribution of the Markov chain defined by the rates $(\tilde{a}_{Z_k,Z_l})_{k,l}$, whenever there exists. We consider that knowing the state $Z = Z_k$, (X, E) follows a PDMP system of the form (1.15) with constant burst rate functions $k_{on,i}(X_{eq,Z_k})$ and $k_{off,i}$. We have the following theorem:

Proposition 10. *Under the hypothesis that the Markov chain defined by the rates $(\tilde{a}_{Z_k,Z_l})_{k,l}$ converges to a stationary distribution, the stationary distribution $\hat{u}$ of the phenomenological model on (X, E, Z) with transition rates on Z defined by (2.1) has the density:*

$$\hat{u}(e,x,z) = \mu_Z(z) \prod_i Beta\left(\frac{k_{on,i}(X_{eq,z})}{d_i}, \frac{k_{off,i}}{d_i}\right)(x) \times \frac{x_i(1-x_i)}{|e_i - x_i|}. \tag{2.2}$$

Proof. We can write the joint probability density $u(t,e,x,z)$ of (E_t, X_t, Z_t) as a 2^n-dimensional vector $u(t,x,z) = (u_e(t,x,z))_{e \in P_E} \in \mathbb{R}^{2^n}$. Using similar arguments that for justifying (1.18), the master equation on u can be written:

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t,x,z) = -\sum_{i=1}^{n} \partial_{x_i}\left(F_i(x)u(t,x,z)\right) + \sum_{i=1}^{n} K_i(x)u(t,x,z) + \sum_{k=1}^{N_Z} a_{Z_k,z}(x)u(t,x,Z_k) - a_{z,Z_k}(x)u(t,x,z).$$

Taking for all $e \in P_E$ the density (2.2), we obtain (see [36] for the details) that

$$\sum_{i=1}^{n} \frac{\mathrm{d}}{\mathrm{d}x_i}\left(F_i(x)\hat{u}(x,z)\right) - \sum_{i=1}^{n} K_i(x)\hat{u}(x,z) = 0,$$

and

$$\left[\sum_{k=1}^{N_Z} a_{Z_k,z}(x)\hat{u}(x,Z_k) - a_{z,Z_k}(x)\hat{u}(x,z)\right]_e = \prod_i \frac{x_i(1-x_i)}{|e_i - x_i|} \times \sum_{k=1}^{N_Z} \left(\tilde{a}_{Z_k,z}\mu_Z(Z_k) - \tilde{a}_{z,Z_k}\mu_Z(z)\right).$$

Thus, if μ_Z is the stationary distribution of the Markov chain defined by the transitions rates $\tilde{a}_{Z_k,Z_l}$, the term on the right hand side is clearly equal to 0. □

Taking the marginal on X of this distribution, we then obtain the Beta mixture of the form (33). Interestingly, the modification (2.1) of the transition rates, with respect to the ones computed by the formula (17), tends to force the jumps between cellular states when the proteins state is far from the basins associated to the actual cellular state. Intuitively, this correction thus improves the coupling between the process defined by the indicator function on the basins (seen as cell types) when (E, X) follows the PDMP process associated to a given GRN, and the marginal on the random variable Z of the phenomenological process associated to this GRN.

Part II

Inference and simulation of gene regulatory networks.

Chapter 3

Method for reverse-engineering a mechanistic model and application to simulated datasets. Article published in In Silico Biology.

We have seen in Chapter 2 that it was possible to reduce the PDMP model (1.15) into a discrete coarse-grained model on the cellular types, and that this reduction motivated the construction of an approximate phenomenological model associated to a GRN, able to reproduce the main dynamics of the process while having an explicit stationary distribution.
In this chapter, we use these results for developing an algorithm which aims to infer a most-likely GRN from a time-course serie of scRNA-seq data. The main idea consists in using the Hamiltonian function found in Chapter 2 associated to a given GRN. To introduce the principle of the method, let us consider that we observe a dataset $X \in \mathbb{R}^{C\times n}$ containing the expression of proteins of C cells for n genes. We have seen in Chapter 2 that under the hypothesis that we can estimate from this dataset the stationary distribution $\hat{u}$ of the underlying PDMP process modeling the differentiation of cells, denoting V the potential associated to this process, we should have the relation

$$-\ln(\hat{u}) = V + O(\varepsilon),$$

where ε characterizes the ratio between promoters and proteins dynamics. Recalling that H^θ denotes the Hamiltonian of the PDMP process driven by the GRN θ, we should then obtain the relation

$$H^\theta(\cdot, \nabla - \ln(\hat{u}(\cdot))) = 0,$$

uniformly on the gene expression space, provided that $\hat{u}$ is a good estimation of the true stationary distribution and that ε is small enough. We recall that H^θ has an explicit form depending on the GRN and some other parameters of the PDMP model (which is given in Chapter 2, formula (18), when the burst rate functions k_{on} are parameterized by θ). Thus, knowing $\hat{u}$, we could estimate the GRN by solving a problem of the form:

$$\theta^* = \min_\theta \int_X H^\theta(x, \nabla - \ln(\hat{u}(x)))dx.$$

As $\hat{u}$ is supposed to be close to a Beta mixture, thanks to the phenomenological model, it should be possible to estimate it from the observations using standard statistical tool like MCMC methods or EM algorithms. This method then seems applicable in practice.

Based on this reasoning, we developed an algorithm that we presented in an article published in the journal *In Silico Biology* [96]. In this work, we use the bursty model (1.17) instead of the PDMP

model (1.15). Indeed, its stationary distribution is close to a mixture of Gamma distributions instead of Beta distributions, which are more convenient to use than Beta distributions in two ways:

- They are not bounded from above, which implies that it is not necessary to assume any maximal value of expression. Indeed, for the PDMP model (1.15), there exists a maximal protein level for every gene i, which is equal to the ratio s_i/d_i. However, these parameters are unknown and estimating this ratio from the data involves risks because there is no guarantee that this maximum level is reached by a cell during the experiment. Thus, it is convenient to not make this assumption for the bursty model (in practice we will see that we still need to assume that for each gene, there is a cell which reaches the maximal burst frequency during the experiment, but this is a weaker assumption);
- When dealing with integer-valued *count* data, as it is the case for scRNA-seq datasets, we can consider that the number of mRNAs expressed by each gene correspond to a Poisson distribution whose mean is the value corresponding to the continuous mechanistic process modeling gene expression [83]. The marginal distributions characterizing the observations can then be considered close to Gamma-Poisson distributions (or mixture of them) when we consider the bursty model (1.17), which correspond to Negative-Binomial distributions. If we considered the PDMP model (1.15), we would obtain Beta-Poisson distributions, which do not correspond to any convenient probability law.

We then extended the method described previously in the realistic case where mRNAs are observed rather than proteins, and proposed an other heuristic extension of the method to the case where we have access to time-stamped datasets.
The resulting algorithm is called CARDAMOM (Cell type Analysis from scRna-seq Data achieved from a Mixture MOdel). It consists in two steps: to characterize in a first time the metastable parameters associated to the coarse-grained model which describes the best the data, and to solve in a second time a serie of regression problems aiming to link these parameters to a most-likely GRN. The simplicity of the regression step is particularly remarkable, and appears close to the learning of a neural network. We show its efficiency on reconstructing a GRN for two *in silico* generated datasets. Such inference method based on a mechanistic model has the great advantage to make quantitative predictions that can then be easily tested by simulating the model, which will be the subject of the next chapter.

Note that the method developed for the first step, that returns the metastable parameters describing the data, has been used in another paper in order to find the most-likely number of metastable basins associated to each gene in four datasets. By analyzing the molecular proximity between reverting and undifferentiated cells in terms of their number of metastable basins, this contributed to demonstrate the biological hypothesis that differentiating cells (chicken erythrocytic progenitors (T2EC)) retain for 24 hours the ability to self-renew when transferred back in self-renewal conditions, in an article accepted for publication in BMC Biology [110].

In Silico Biology 14 (2021) 89–113
DOI 10.3233/ISB-210226
IOS Press

Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics

Elias Ventre[a,b,c,*]
[a]*Laboratory of Biology and Modelling of the Cell, ENS de Lyon, CNRS UMR 5239, Lyon, France*
[b]*Inria Center Grenoble Rhone-Alpes, Equipe Dracula, Villeurbanne, France*
[c]*Institut Camille Jordan, Université Claude Bernard Lyon 1, CNRS UMR 5208, Villeurbanne, France*

Abstract. Differentiation can be modeled at the single cell level as a stochastic process resulting from the dynamical functioning of an underlying Gene Regulatory Network (GRN), driving stem or progenitor cells to one or many differentiated cell types. Metastability seems inherent to differentiation process as a consequence of the limited number of cell types. Moreover, mRNA is known to be generally produced by bursts, which can give rise to highly variable non-Gaussian behavior, making the estimation of a GRN from transcriptional profiles challenging. In this article, we present CARDAMOM (Cell type Analysis from scRna-seq Data achieved from a Mixture MOdel), a new algorithm for inferring a GRN from timestamped scRNA-seq data, which crucially exploits these notions of metastability and transcriptional bursting. We show that such inference can be seen as the successive resolution of as many regression problem as timepoints, after a preliminary clustering of the whole set of cells with regards to their associated bursts frequency. We demonstrate the ability of CARDAMOM to infer a reliable GRN from in silico expression datasets, with good computational speed. To the best of our knowledge, this is the first description of a method which uses the concept of metastability for performing GRN inference.

Keywords: Single cell, gene regulation network, inference, metastability, transcriptional bursting, machine learning

1. Introduction

Differentiation is the process whereby a cell acquires a specific phenotype, by differential gene expression as a function of time. Measuring how gene expression changes as differentiation proceeds is therefore of essence to understand differentiation. Advances in measurement technologies now allow to obtain gene expression levels at the single cell level. It offers a much more accurate view than population-based measurements, that has been obscured by mean population-based averaging [9, 25]. It has among other things established that there is a high cell-to-cell variability in gene expression, and that this variability has to be taken into account when examining a differentiation process at the single-cell level [3, 13, 28, 30, 33, 41, 46, 47].

A popular vision of the cellular evolution during differentiation, introduced by Waddington in [52], is to compare cells to marbles following probabilistic trajectories, as they roll through a developmental landscape of ridges and valleys. The landscape is generally described by the graph of a potential function $V : \Omega \to \mathbb{R}$, where Ω denotes the space where these trajectories evolve, and is called the gene expression space. In that space, a cell can be described by a vector, where each coordinate represents the expression products of a gene [17, 31].

A cell has theoretically as many states as the combination of proteins and mRNAs quantity possibly associated to each gene, which is potentially huge [7]. But metastability, that is the coexistence of multiple stable states, seems inherent to cell differentiation

*Corresponding author: Elias Ventre. Email: elias.ventre@ens-lyon.fr.

processes as evidenced by limited number of existing cellular phenotypes [4, 32]. Since [20] and [18], many authors have identified cell types with the basins of attraction of a dynamical system modeling the differentiation process. In that context, noise in gene expression, or cell to cell heterogeneity, appears to be closely related to the transition between these cell types [12]. This provides a rationale for coarse-graining stochastic models of gene expression into reduced processes on a limited number of metastable basins, seen as cell types. Such reduction has been studied mostly in the context of stochastic diffusion [53, 54, 57], but also for stochastic hybrid systems [23]. This last case is particularly interesting because the basins can be classified regarding to the modes associated to the jumps frequency of the discrete variable, leading to local approximations of the potential function V describing the landscape [51].

This landscape is often regarded to be shaped by an underlying gene regulatory network (GRN), which appears as a powerful abstraction for describing interactions between genes through their proteins production. The construction of GRNs from literature being a very time-consuming and labor intensive process, and sometimes impossible due to the limitation of our current knowledge, their automated reconstruction from large datasets has become a classic task in systems biology [44]. This task is notoriously difficult, in particular when dealing with single-cell transcriptomics, the bursty synthesis of mRNAs [35, 58] giving rise to highly variable and non-Gaussian expression data [26]. The methods that are used cover a wide range of statistical and modeling tools [1], including the analysis of stochastic models of gene expression [5, 16]. In the latter case, expression datasets are identified to independent samples of the time-varying distribution describing the process. GRN inference can then be seen as the reconstruction of the most-likely GRN from a set of partial observations of independent realizations of the model.

To our knowledge, the use of metastability for performing such reverse engineering has not been studied yet. The main contribution of this article is to derive an efficient algorithm for linking the landscape analysis of a mechanistic model of gene expression using metastability, to the most-likely associated GRN parameters. In Section 2, we are going to present a mechanistic model of gene expression developed in [16], which describes single-cell dynamics associated to a GRN with transcriptional bursting. With a methodology similar to the one developed in [51], we perform its reduction into a discrete coarse-grained model on a limited number of metastable basins. We deduce from this reduction an approximation of the time-dependent proteins distribution of the original model, presented in Section 3, which appears as a mixture of Gamma distributions. We develop in Section 4 a statistical method for linking the parameters of such mixture to the GRN parameters of the model. In Section 5, we extend this method for estimating GRN parameters from scRNA-seq timestamped datasets. We show in Section 6 the accuracy of the method for *in silico* datasets simulated from the mechanistic model with various networks of different sizes. Finally, we discuss more precisely in Section 7 the interpretation of the method in terms of landscape, its applicability to real datasets, and we highlight its similarity with a machine learning approach.

We draw the reader's attention to the fact that the problem of inferring a GRN using a mechanistic model with transcriptional bursting has been recently elsewhere using distinct mathematical tools [15]. We are currently setting up a collaborative effort for benchmarking both algorithms on in silico generated data as well as on real datasets from the literature. This benchmarking is therefore beyond the scope of the present article and will be exposed elsewhere.

2. GRN model and reduction

2.1. Mechanistic model

The model which is used throughout this article has been introduced in [16]. It is based on a hybrid version of the well-established two-state model of gene expression [21, 38], where a gene is described by the state of a promoter, which can be $\{on, off\}$. If the promoter is on, mRNAs are transcribed at a rate s_0, which are then translated into proteins at a rate s_1. Degradation of both mRNAs and proteins occurs respectively at a rate d_0 and d_1. The transitions between the states on and off occur at exponential times of rates k_{on} and k_{off}. We consider the so-called bursty regime of this model, when $k_{on} \ll k_{off}$, which corresponds to the experimentally observed situation where active periods are short but characterised by a high transcription rate, thereby generating bursts of mRNA [34, 43, 50]. We describe the random times at which these bursts occur by an exponential law of parameter k_{on}, and their random intensity by an exponential law of parameter k_{off}/s_0 (see Figure 1). This model is compatible with real single-cell data, as

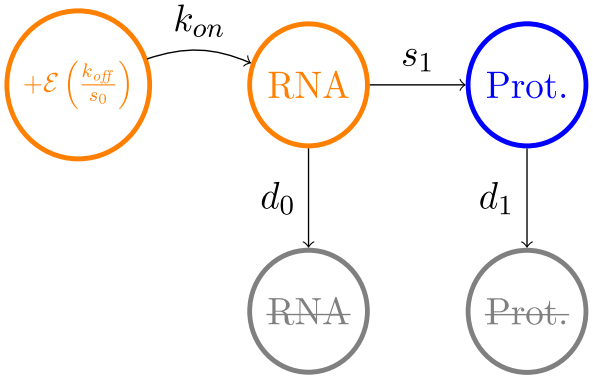


Fig. 1. Approximation of the two-states model of gene expression in the bursty regime.

mRNAs quantity at the steady state follows a Gamma distribution, which is known to describe accurately continuous single-cell data [2].

Neglecting the molecular noise associated to mRNA and protein quantities, we obtain the following mathematical description of the model:

$$\begin{cases} M(t) \xrightarrow{k_{on}} M(t) + \mathcal{E}\left(\frac{k_{off}}{s_0}\right), \\ M'(t) = -d_0 M(t), \\ P'(t) = s_1 M(t) - d_1 P(t). \end{cases} \tag{1}$$

where $M(t)$ and $P(t)$ denote respectively the mRNA and protein concentration at time t and $\mathcal{E}\left(\frac{k_{off}}{s_0}\right)$ is an exponential law of mean $\frac{s_0}{k_{off}}$. The key idea for studying a GRN is to embed this model into a network. Denoting the number of genes by n, the vector $(M,\ P)$ describing the process is then of dimension $2n$. The burst rates for each gene i are characterized by two gene-specific functions $k^{\theta}_{on,i}$ and $k_{off,i}$. For the sake of simplicity, we consider that $k_{off,i}$ does not depend on the protein level (*i.e* that proteins do not affect the quantity of mRNAs which are transcribed during a burst). To take into account the interactions between the genes, we consider that for all $i = 1, \cdots, n$, $k^{\theta}_{on,i}$ is a function which depends on the full vector P via the GRN, represented by a squared matrix θ of size n. We call these functions the bursts rate functions in the following. We define for all $i = 1, \cdots, n$:

$$k^{\theta}_{on,i}(P) = k_{0,i} + (k_{1,i} - k_{0,i})\sigma^{\theta}_i(P), \tag{2}$$

where $\sigma^{\theta}_i(P) = \left(1 + \exp\left(-\beta_i - \sum_{j=1}^{n} \theta_{ij} P_j\right)\right)^{-1}$. The parameter β_i represents the basal activity of gene i, and each parameter θ_{ij} encodes the interaction $j \to i$. Every function $k^{\theta}_{on,i}$ is then comprised between two positive constants $k_{0,i} < k_{1,i}$, and $\partial_{P_j} k^{\theta}_{on,i}$ has the sign of θ_{ij}. This sigmoidal form for the bursts rate functions can be interpreted as a simplification of the mechanistic form used in [5, 16].

We point out that assuming that this model is able to reproduce real single-cell datasets, we make the underlying hypothesis, which is implicit in most GRN inference methods, that, contrary to its state, the GRN structure is not modified under the action of some hidden variables. One should not that our interaction model is an approximation of the underlying biochemical cascade reactions, and that the genes we modeled need not to be transcription factors. This is possible thanks to the use of our mechanistic model which integrates the notion of timescale separation [16]. It assumes that every biochemical reaction such as metabolic changes, nuclear translocations or post-translational modifications are faster than gene expression dynamics and that they can be abstracted in the interaction between 2 genes [5]. To explicitly model slow changes like some epigenetic changes by allowing the structure of the GRN to change during differentiation, would definitely add realism but at the cost of a much higher complexity: the number of additional unobserved parameters would dramatically increase the problems of identifiability, making a reverse-engineering method out of reach.

2.2. *Simplification in the fast transcription regime*

In line with several experiments [2, 22], we consider that mRNA bursts are fast in regard to protein dynamics, *i.e* $d_{0,i} \gg d_{1,i}$ with $\frac{k_{1,i}}{d_{0,i}}$ fixed. The correlation between mRNAs and proteins produced by a gene i is then very small, and the model can be reduced by removing mRNA and making proteins directly depend on the burst. We obtain a simplified network model in which only proteins are described:

$$\begin{cases} P_i(t) \xrightarrow{k^{\theta}_{on,i}(P_i(t))} P_i(t) + \mathcal{E}(c_i), \\ P_i'(t) = -d_{1,i} P_i(t), \end{cases} \tag{3}$$

where we define $c_i = \frac{k_{off,i} d_{0,i}}{s_{0,i} s_{1,i}}$. The gene expression space Ω is then the set of possible values for the vector P, which is $\mathbb{R}^{n+}$. The related master equation, characterizing the time-dependent distribution of P, turns out to be integro- differential (see Appendix B).

2.3. Metastability and reduction

We now perform a reduction of the model 3 into a coarse-grained model on a limited number of metastable basins, providing an approximate landscape of the differentiation process. For this sake, we introduce a typical time scale $\bar{k}$ for the rates of promoters activation $k^{\theta}_{on,i}$, and a typical time scale $\bar{d}$ for the rates of proteins degradation. Then, we define the scaling factor $\varepsilon = \frac{\bar{d}}{\bar{k}}$ which characterizes the difference in dynamics between two processes: 1. gene bursting dynamics and 2. protein dynamics. It is generally considered that promoter switches are fast with respect to protein dynamics, *i.e* that $\varepsilon \ll 1$, at least for eukaryotes [48].

In that context, we can approximate the conditional expectation of the bursts of proteins associated to a gene i knowing the proteins vector P, that we denote $\rho_i(P)$, by its quasistationary approximation $\overline{\rho_i}(P) = \frac{k^{\theta}_{on,i}(P)}{c_i}$. The model (3) can therefore be coarsely approximated by a system of ordinary differential equations:

$$\forall i = 1, \cdots, n : \; P_i'(t) = \frac{k_{on,i}(P(t))}{c_i} - d_{1,i} P_i(t). \tag{4}$$

Intuitively, these trajectories correspond to the mean behaviour of a cell in the weak noise limit, i.e when bursts occur much faster than proteins concentration changes. As shown in [11] (in the general case where promoters are explicitly included in the model), for any $T < \infty$, a random path $(X^{\varepsilon}(t))_{0 \leq t \leq T}$ converges in probability to a trajectory $(x(t))_{0 \leq t \leq T}$ solution of the system (4) when $\varepsilon \to 0$.

Assuming that the dynamical system (4) has no limit cycles or more complicated orbits, the gene expression space can be then decomposed in a set of basins of attraction $Z = \{Z_1, \cdots, Z_m\}$, respectively associated to m stable solutions of:

$$\forall i = 1, \cdots, n : \; \frac{k_{on,i}(P)}{c_i d_{1,i}} - P_i = 0, \tag{5}$$

which are called the attractors of the process. Without noise, the fate of a cell is fully characterized by its initial state x_0, as it converges to the attractor of the basin of attraction it belongs to, which is a single point by assumption. However, noise can modify the deterministic trajectories in at least two ways. First, in short times, a stochastic trajectory can deviate significantly from the deterministic one. In long time, stochastic dynamics can even push the trajectory out of its basin of attraction to another one, changing radically the fate of the cell in a way that cannot be catched by the deterministic limit. We illustrate in Figure 2 this situation for a toggle-switch network of two genes, where the scaling factor ε determines the observation of random transitions between two basins of attraction in a given time.

Adopting the paradigm of metastability referred in the introduction, we identify the basins of attraction associated to the equilibrium of the deterministic system (4) to cell types [18]. A cell type then corresponds to a metastable sub-region of the gene expression space, and the process can be coarsely reduced to a new (Markovian) discrete process on the cell types. These cell types represent the potential wells in the developmental landscape of differentiation associated to the model (3) [49], the centers of which are the attractors solutions of the system (5). To characterize more precisely the landscape, it would remain to describe:

1. The energetic barrier separating the cell types;
2. The curvature of the potential wells.

Point 1. corresponds to the transition rates of the coarse-grained model on the cell types, which are known to be very difficult to link analytically to a GRN [51]. They generally depends on the values of the stationary distribution $\hat{u}^{\theta}$ of proteins on the saddle point of the system (4), which are located on areas of the gene expression space where the probability to find a cell is weak. As discussed in Section 7.2 with more details, this feature does not seem to be exploitable in the context of GRN inference. Point 2. can be described by the behaviour of the potential function $V^{\theta} = -\ln\left(\hat{u}^{\theta}\right)$ in the neighborhood of the attractors. This feature seems more accessible because cells are likely to be measured around the attractors in the gene expression space. We are now going to develop an heuristic reasoning for approximating the function $\hat{u}^{\theta}$ when the bursts rate functions are of the form (2).

3. From GRN to mixture approximation

3.1. Mixture approximation of the proteins distribution

On one side, the transitions of the coarse-grained model described in Section 2.3 happen at a time scale which is expected to be of the order of $e^{\frac{C}{\varepsilon}}$, where C is an unknown constant depending on the

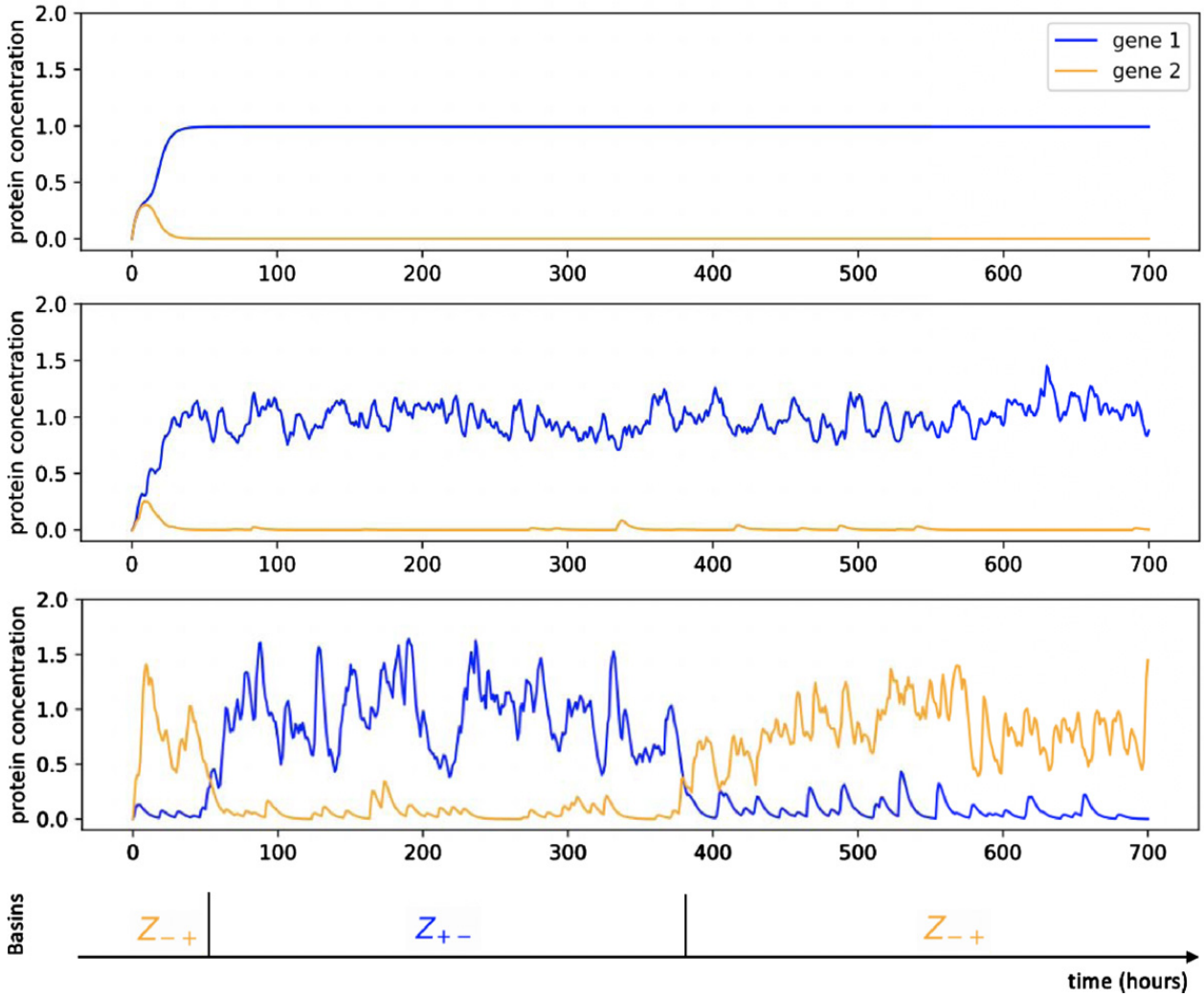


Fig. 2. Example of trajectories associated to the symmetric toggle-switch network described in Table 1, for different values of ε: from top to bottom, $\varepsilon = 0$, $\varepsilon = 1/50$ and $\varepsilon = 1/10$. For $\varepsilon = 1/10$, we observe stochastic transitions between the two metastable basins denoted Z_{+-} and8 Z_{-+}.

basins (owing to a Large deviations principle studied in [51]). For small ε, such transitions are then generally rare events, and it can be considered that the process spends in each basin a time long enough to equilibrate inside, *i.e* that a cell reaches between each transition its quasistationary distribution within a basin. On the other side, the sigmoidal form (2) for the functions $k^{\theta}_{on,i}$ implies that these functions must not vary significantly within a basin: a rough approximation could lead to identify the bursts rate in each basin by its dominant rate inside, corresponding to the value of the function $k_{on,i}$ on the attractor P_z. For any gene $i = 1, \cdots, n$ and any basin $z \in Z$, we can then approximate:

$$\forall P \in z : \; k^{\theta}_{on,i}(P) \approx k^{\theta}_{on,i}(P_z).$$

This implies that the quasistationary distribution of a cell within a basin can be approximated by the stationary distribution of the model when the burst rates are constant. The marginal distribution of each gene then appears as a Gamma distributions, which can be shown to be the unique solution of the stationary master equation of the model in one dimension when k_{on} is constant [24]. We obtain the following approximation for the quasistationary distribution $\hat{u}^{\theta}_z$ within each basin $z \in Z$ associated to the network θ:

$$\hat{u}^{\theta}_z \approx \prod_{i=1}^{n} \text{Gamma}\left(\frac{k^{\theta}_{on,i}(P_z)}{d_{1,i}}, c_i\right),$$

Finally, we can approximate the stationary distribution of the process associated to a given GRN $\hat{u}^{\theta}$ by a mixture of Gamma distributions. Denoting for any $z \in Z$, $i = 1, \cdots, n$, $k_{z,i} = k^{\theta}_{on,i}(P_z)$, we obtain

$$\hat{u}^{\theta} \approx \hat{u}_{\varepsilon} = \sum_{z \in Z} \mu(z) \prod_{i=1}^{n} \text{Gamma}\left(\frac{k_{z,i}}{d_{1,i}}, c_i\right), \quad (6)$$

where μ is the probability vector on the basins at the steady-state, and we recall that ε depends on the values of the vectors d_1 and k_z. In some sense, this approximation consists in reducing the dependence between genes resulting from the GRN to the coexistence and relative weight of different basins corresponding to the different possible modes of promoters frequency.

The heuristic analysis presented above then states that when the functions $k_{on,i}$ are close to constant functions inside the basins and that the scaling factor ε is small, the mixture approximation (6) is a good approximation of the distribution $\hat{u}^\theta$. We remark that this is straightforward when the scaling factor ε is close to 0. Indeed, any GRN θ is associated for any value of ε to an unknown stationary distribution, that we denote $\hat{u}^\theta_\varepsilon$, which converges to a sum of Dirac on the attractors when $\varepsilon \to 0$ [36]. Then we see that the Gamma mixture $\hat{u}_\varepsilon$ associated to the same ε converges to the same sum of Dirac when $\varepsilon \to 0$, as the mean of each Gamma does not depend on ε while the variance is proportional to ε: we have $|\hat{u}^\theta_\varepsilon - \hat{u}_\varepsilon| \to 0$ as $\varepsilon \to 0$. When the noise ε is not negligible, the problem is more complex and finding a theoretical bound for the quantity $|\hat{u}^\theta_\varepsilon - \hat{u}_\varepsilon|$, depending on ε, is beyond the scope of this paper. Nevertheless, we see in Figure 3 that the Wasserstein distance between the empirical distribution associated to a proteins dataset simulated from the mechanistic model (3) and the Gamma mixture distribution (6) is much smaller than the distance between the same dataset and a mixture of normal distribution (fitted with a Gaussian Mixture Model). This let us think that the Gamma mixture approximation is indeed close to the true distribution even when the weak noise limit assumption is not realistic (and that the distribution is then far from a sum of Diracs).

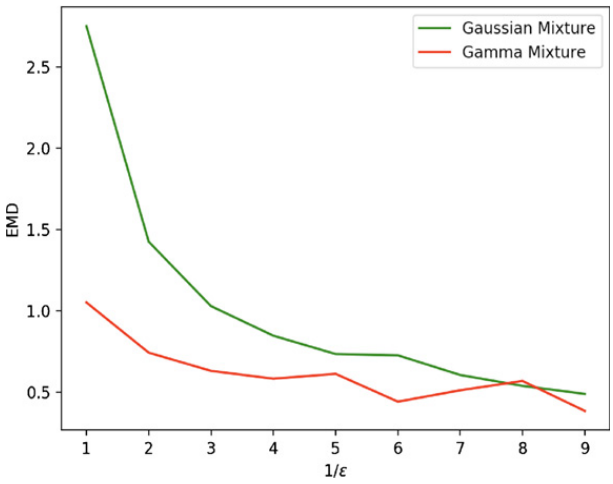


Fig. 3. Evolution in function of $1/\varepsilon$ of the Wasserstein distances between the empirical distribution associated to a set of 2000 cells, simulated with the model (3), and both the Gamma mixture approximation (6) (in red) and a GMM approximation (in green), for the toggle-switch network described in Table 1.

Table 1
Description of the parameters of the symmetric two-dimensional toggle-switch used for the illustrations of Sections 2, 3 and Appendix C, and the network for the inference in Section 6.2 which consists in two such toggle-switch functioning in parallel

(i,j) / param.	$k_{0,i}$	$k_{1,i}$	$d_{1,i}$	c_i	$\theta_{i,i}$	$\theta_{i,j}$
(1,2)	0, 1	2	0, 2	10	5	−7
(2,1)	0, 1	2	0, 2	10	5	−7

3.2. Linking a GRN to mixture parameters

Building an analytical link between the GRN and the mixture parameters is generally out of range. Indeed, even if it was possible to solve explicitly the equation (5) defining the equilibrias associated to a GRN, it would remain challenging to study their stability in order to identify the attractors. Moreover, the probability vector μ is obviously linked to the transition rates of the coarse-grained model, which we recall to be very difficult to estimate from a GRN [51]. However, these parameters can be obtained with a simple numerical method, which consists in sampling a collection of random paths in the gene expression space: the distribution of their final position after a long time approximates the stationary distribution on proteins. We can then use these final positions as starting points for simulating the deterministic trajectories, given by the system (4), with an ODE solver: each of them converges to one of the stable equilibrium points. This method allows to obtain all the stable equilibria corresponding to sufficiently deep potential wells (see Figure 4(A)). Possible other potential wells can be omitted because they should correspond to basins that the process has very low probability of visiting, which do not impact significantly the coarse-grained Markov model. Repeating this operation for thousands of cells, we can find a vector μ describing the ratio of cells belonging to each basin (see Figures 4(B) and 4(C)). When the number and length of the simulations are large enough, the vector μ should be a good approximation of the stationary measure on the basins.

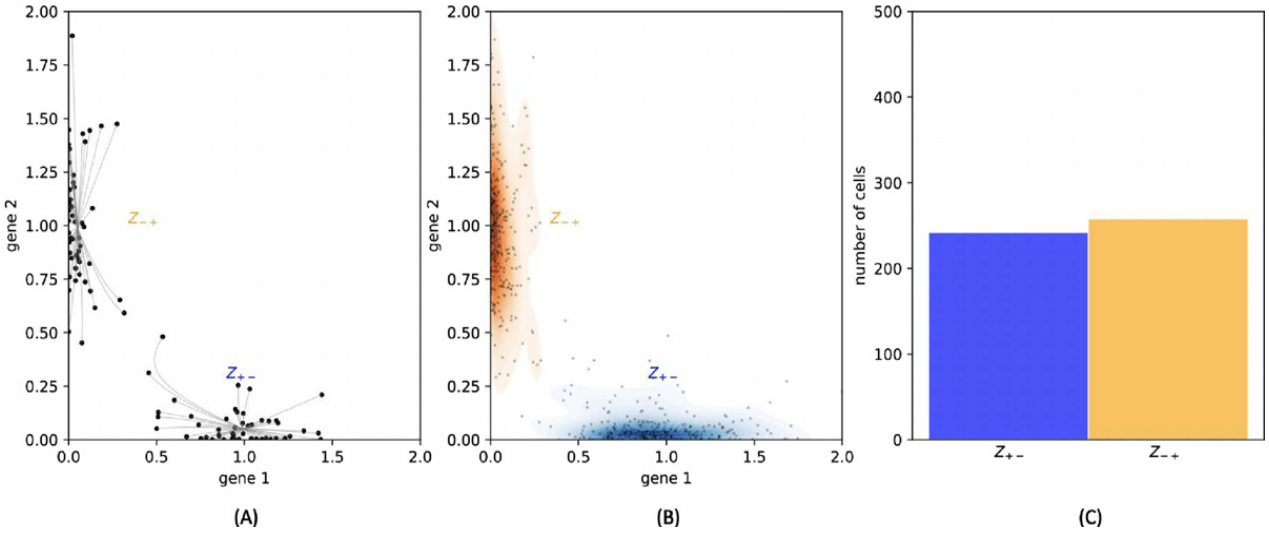


Fig. 4. (A): 100 cells are plotted under the stationary distribution. The relaxation trajectories allow to link every cell to its associated attractor. (B): 500 cells are plotted under the stationary distribution. They are then classified depending on their attractor, and this figure sketches the kernel density estimation of proteins within each basin. (C): The ratio of cells that are found within each basin gives an estimation of the stationary distribution on the basins.

3.3. Mixture approximation of time-varying distribution

In the mixture approximation, the marginal on proteins of the stationary distribution of a single cell is characterized by a hidden Markov model: in each basin $z \in Z$, which corresponds to the hidden variable, the vector P is randomly chosen under the quasistationary distribution $\hat{u}_z$ of $X \mid Z = z$. Therefore, the mixture distribution can also be used as a proxy for the time-varying distributions of the bursty model (3). Denoting μ_t the distribution on the basins at any time t, and considering that a cell reaches almost immediately its quasistationary distribution after entering in a basin, we obtain the approximation:

$$u_t^\theta \approx \sum_{z \in Z} \mu_t(z) \prod_{i=1}^{n} \text{Gamma}\left(\frac{k_{z,i}}{d_{1,i}}, c_i\right).$$

In that case, the only time-dependent parameters are the coordinates of the vector $\mu_t \in [0, 1]^m$ where m is the number of basins, and $\mu_t(z) = \mu(z)$ if t is such that the stationary distribution is reached.

We remark that with the method described in Figure 4, it would be possible to miss some basins that are important for describing the network, but which would not appear in the stationary distribution. It is expected to be a common situation for networks showing complex behaviours, as feedback loops or unbalanced branching structure, because the probability after a long time may be too weak for some basins to be visited, while playing an important role in the process beforehand. For such networks, the method should then be applied on a series of timepoints, and the union of all the basins identified at every timepoint should be considered (fixing the value $\mu_t(z) = 0$ for basins z that are not observed at time t). In that point of view, the basins appearing in the stationary distribution can be considered as (almost) absorbing states of the coarse-grained model.

Altogether the reduction and methods described above establish a formal basis for the definition of a simplified epigenetic landscape given a GRN, under the form of a mixture model. As the mixture parameters seem possible to obtain from a single-cell dataset (see Section 5.2), it would be fruitful to use the same formalism to assess the inverse problem of inferring the most likely GRN, given an (experimentally-determined) cell distribution in the gene expression space, a notoriously difficult task [16, 39].

4. Method for inferring a GRN from a mixture distribution

In this section, we discuss the problem of identifying the most-likely GRN which is associated to an approximate landscape described by the Gamma mixture distribution of the form (6). We identify such landscape to the set of parameters:

$$\alpha = \left\{\mu(z), \frac{k_{z,i}}{d_{1,i}}, c_i\right\}_{z \in Z, i=1,\cdots,n}. \qquad (7)$$

The preliminary problem of finding the most-likely vector α from a single-cell dataset using classical tools of statistical analysis will be developed in the presentation of the final algorithm (see Section 5.2.1).

4.1. Linking the GRN and the attractors

The main idea of the method consists in using the fact that all the attractors P_z of the bursty model verify the equation (5), *i.e* that for all $i = 1, \cdots, n$,: $k^{\theta}_{on,i}(P_z) = c_i d_{1,i} P_{z,i}$.

Thus, from the definition of $k_{z,i} = k^{\theta}_{on,i}(P_z)$, we have the relation:

$$\forall i = 1, \cdots, n,\ \forall z \in Z :\ k^{\theta}_{on,i}\left(\frac{k_z}{cd}\right) = k_{z,i}. \quad (8)$$

Considering that α is known and that we have to determine θ, we obtain for every gene i a system of m equations and n unknowns parameters corresponding to the i^{th} line of the matrix θ. This simple strategy may be efficient provided that there is enough basins in regards to the number of genes of interest. This is in particular the case for the toggle-switch network described in Table 1, for which its has been observed in Figure 4 that there are exactly two attractors.

4.2. GRN inference from a Gamma mixture as a minimization problem

GRN inference from a Gamma mixture is formally equivalent to identifying an adapted function R such that the "real" GRN θ^* associated to a set of mixture parameters α would be defined by

$$\theta^* = \underset{\theta \in M_n(\mathbb{R})}{\arg\min}\, R(\theta, \alpha).$$

It is nevertheless important to remark that they may be no function R such that this equality is hold for any GRN, as a high (unknown) number of GRNs could be associated to the same mixture distribution (6). This problem is not specific to our method, and GRN inference is known to be generally a non-identifiable problem [5].

Regarding to the analysis which has been made in Section 4.1, a natural candidate for this function is

$$R(\theta, \alpha) = \sum_{z \in Z} \sum_{i=1}^{n} \left(k^{\theta}_{on,i}\left(\frac{k_z}{cd}\right) - k_{z,i} \right)^2, \quad (9)$$

to which it may be added an adapted penalization (see Section 5.2). Taking into account the probability vector μ is discussed at the end of Section 5.2.2, and stems for a generalization of the method presented in this section, developed in Appendix C.

5. Method for inferring a GRN from timestamped scRNA-seq data

In this section, we use the analysis provided in Sections 3 and 4 for developing a numerical method able to infer a GRN from scRNA-seq data, which represents the most available single-cell data at present. From a statistical point of view, scRNA-seq data gives access to the joint probability distribution of mRNAs levels associated to a given set of genes in a set of individual cells independently. Importantly, measurement techniques usually involves the physical destruction of the cell. Then, when measurements are made at several time points, for example to study convergence to a possibly new steady state after applying a perturbation to the system, we assume that the data correspond to independent samples of the marginal on mRNA of the time-varying distribution associated to the mechanistic model (1).

5.1. Simplified statistical model for the data

Along with the system of n coupled systems of the form (3) describing proteins dynamics, as presented in Section 2.2, we obtain when $\varepsilon \ll 1$, *i.e* when the bursts are frequent in regard to proteins dynamics, a quasi steady-state approximation for the conditional distribution of M given P. Under this approximation, mRNAs levels M_i are independent conditionally to the protein vector, and follow Gamma distributions depending on P:

$$M_i \mid P \sim \text{Gamma}\left(\frac{k_{on,i}(P)}{d_{0,i}}, \frac{k_{off,i}}{s_{0,i}}\right).$$

Combining this statistical model of gene expression with a Poisson model, which is claimed in [45] to adequately describe the measurement process, we then obtain the following model for a set of observed single-cell mRNA transcriptomic data:

$$M_i \mid P \sim \text{NB}\left(\frac{k_{on,i}(P)}{d_{0,i}}, \frac{k_{off,i}}{k_{off,i} + s_{0,i}}\right).$$

Here, NB denotes the negative binomial distribution: $\forall k \in \mathbb{N} :\ \text{NB}(k, a, b) = \frac{\Gamma(k+a)}{\Gamma(a)k!} b^k (1-b)^a$.

Contrary to what has been done in [15], where the inference strategy consists in treating proteins level as latent variables and using the law on mRNAs knowing

proteins as a statistical likelihood for the data, we are going to use the analysis provided in Section 3 to directly estimate the set of mixture parameters α from the data without the need of proteins quantity. The main idea is the following. On one side, the only information on the proteins that can be obtained is contained in the value of $k_{on,i}(P)$. On the other side, proteins being known to be less noisy than mRNAs, a vector P characterizing a cell in the gene expression space is going to be generally close to one of the attractor. Moreover, the functions $k_{on,i}(P)$ are almost constant within each basin, and in particular around their attractor. Thus, the precise knowledge of the proteins quantity can be considered out of reach from scRNA-seq data, and the best information on proteins we can get from the marginal distribution on mRNAs of a gene i is the set of the modes of bursts frequency normalized by the degradation rate, that is $\frac{k_{z,i}}{d_{0,i}}$.

We finally derive a simplified statistical model, making appear the distribution of mRNA counts as a mixture of negative binomial:

$$M \sim \sum_{z\in Z} \mu(z) \prod_{i=1}^{n} \mathrm{NB}\left(\frac{k_{z,i}}{d_{0,i}}, \frac{k_{off,i}}{k_{off,i}+s_{0,i}}\right). \quad (10)$$

5.2. Inference procedure

We now describe the inference procedure for a set of l timestamped scRNA-seq data $M = (M_{t_1}, \cdots, M_{t_l})$, where each M_{t_k} is a matrix of gene expression containing the n_{t_k} cells $(x^k_1, \cdots, x^k_{n_{t_k}})$. We decompose the algorithm in 2 steps:

1. A clustering step, for identifying the set of parameters α^k_M associated to the data at each timepoint t_k;
2. A regression step, consisting in identifying the most-likely GRN θ associated to the sets $(\alpha^k_M)_{k=1,\cdots,l}$, by successively solving a set of regression problems.

5.2.1. Clustering

From the statistical model (10), the first step consists in inferring for each timepoint t_k the set of parameters

$$\alpha^k_M = \left\{\mu_{t_k}(z), \alpha_z, c\right\}_{z\in Z},$$

where $c = \frac{k_{off}}{k_{off}+s_0} \in \mathbb{R}^n$ and for every $z \in Z$, $\alpha_z = \frac{k_z}{d_0} \in \mathbb{R}^n$. The number of genes being potentially large, a minimization of the maximum likelihood function f on the gene expression matrix $M_{t_k} \in \mathcal{M}_{n_{t_k},n}(\mathbb{N})$, defined by

$$f(M_{t_k}) = \prod_{j=1}^{n_{t_k}} \left(\sum_{z\in Z} \mu_{t_k}(z) \prod_{i=1}^{n} \mathrm{NB}\left(\frac{k_{z,i}}{d_{0,i}}, \frac{k_{off,i}}{k_{off,i}+s_{0,i}}\right)(x^k_j) \right),$$

even using an EM algorithm or an other variational method, would be very uncertain, such algorithms being known to be trapped into local minima [10]. The number of components of the mixture, *i.e* the size of the set Z, remains also unknown. A good method for fitting α^k_M would be to apply a Bayesian procedure like a Reversible-Jump Monte-Carlo Markov Chain (RJMCMC) algorithm [42] to the whole dataset. We propose a slightly different approach, which overcomes the numerical limits of such Bayesian method on the multivariate procedure, consisting in two steps:

1. We first identify the modes of the bursts frequency which are associated to every gene i, independently, for the whole set of cells M. For this sake, we use a RJMCMC algorithm which is inspired from [6]. We obtain the values of $\alpha_{z,i} = \frac{k_{z,i}}{d_{0,i}}$. Each cell $y \in M_{t_k}$, represented by a vector in $\mathbb{N}^n$, can be then associated to a discretized representation of the form: $\tilde{y} = \alpha_z = (\alpha_{z_1}, \cdots, \alpha_{z_n})$, where every z_i characterizes one of the modes associated to the gene i;
2. For each timepoint t_k, we group together the vectors $\tilde{y}$ which are equals. The different groups that are obtained correspond to the modes $(\alpha_z)_{z\in Z}$ observable at t_k. The relative weights associated to these modes provide an estimation of the probability vector μ_{t_k}.

This simplified method, in which the coupling between the different genes is taken into account only in a second stage, is likely to provide more basins than the expected number. However, we will see in Section 5.2.2 that this may be corrected by a simple modification of the cost function used in the regression step. We also discuss in Section 6.5 the fact that the number of modes associated to each gene is generally limited to 2 in the case of the sigmoidal bursts rate functions (2): then each cell $y \in \mathbb{N}^n$ is generally projected in step 1. in a reduced discrete space of dimension 2^n.

5.2.2. Regression

First, in order to reduce the number of parameters and since protein levels P_i are not observed, we can

arbitrarily set:

$$s_{1,i} = \frac{d_{1,i} d_{0,i} k_{off,i}}{k_{1,i} s_{0,i}}, \tag{11}$$

which leads to $c_i = \frac{k_{1,i}}{d_{1,i}}$. Injecting this value in the formula (5), the attractors are then defined for any $z \in Z$ by the formula: $P_z = \frac{k_z}{k_1}$. It is worth noticing that the choice of this scaling may not affect the accuracy of the inference. Indeed, the value of the parameter $s_{1,i}$ only affects the number of proteins P_i that are created when M_i is positive: the value of P_i is then proportional to $s_{1,i}$. As every parameter θ_{ij} affects the dynamics of the process through the value of the product $\theta_{ij} \times P_i$ in the function $k_{on,j}$, considering such scaling for $s_{1,i}$ will only affect the scaling for the interaction coefficients θ_{ij}, but not the GRN topology.

The clustering step only allows to find the values of the vectors $\alpha_z = \frac{k_z}{d_0}$ rather than k_z. This is not a problem since the fraction which characterizes P_z allows to simplify the term d_0, provided that $\frac{k_1}{d_0}$ can be computed. We therefore consider that each gene reaches its minimal and maximal frequency through the timepoints, at least for some cells. Thus, we define for every gene i, $k_{0,i} = \min_z k_{z,i}$ and $k_{1,i} = \max_z k_{z,i}$, and equivalently:

$$\alpha_{0,i} = \min_z \alpha_{z,i},\ \alpha_{1,i} = \max_z \alpha_{z,i}. \tag{12}$$

We finally obtain a characterization of the attractors which is directly accessible from scRNA-seq data:

$$P_z = \frac{\alpha_z}{\alpha_1}. \tag{13}$$

Injecting (13) in the definition of R given by (9), we obtain the following characterization of a cost between a GRN θ and a vector α describing the approximate landscape of the model (1):

$$R(\theta, \alpha) = \sum_{z \in Z} \sum_{i=1}^{n} \left(\sigma_i^\theta \left(\frac{\alpha_z}{\alpha_1} \right) - \frac{\alpha_{z,i}}{\alpha_{1,i}} \right)^2. \tag{14}$$

In order to define a regression problem on the variable θ which may associate a vector α obtained in the clustering step (see Section 5.2.1) to a GRN, we add three modifications to the cost function R given in (14).

First, we solve the problem of the possibly too high number of basins due to the clustering method, that we mentioned in Section 5.2.1, by taking into account the weight of the basins in the cost function. Then, given that we have enough cells that are detected in existing basins, the "false" basins detected in the clustering step, which should not be associated to many cells, would almost not affect the inference.

Second, in line with the idea that missing interactions is preferable to inferring false interactions between genes, we decide to use a LASSO penalization, which is known to enforce the sparsity of the network. We also add a custom penalization to deal with oriented interactions. Indeed, for every pair of nodes $\{i, j\}$ there are two possible interactions with respective parameters θ_{ij} and θ_{ji}, but it is likely that only one is actually present in the true network. Our method is likely to favor symmetric interactions because when an interaction is present in the network (e.g $\theta_{ij} > 0$), then gene j is generally upregulated in the same time than gene i and it is hard to distinguish whether $\theta_{ji} > 0$ or not. Then we want these two interaction parameters to compete each other, such that only one is nonzero after the regression step, unless there is enough evidence in the data that both interactions are present. We obtain the following penalization

$$|\theta - \theta^0| = \sum_{i,j=1}^{n} |\theta_{i,j} - \theta^0_{i,j}| + \frac{1}{2} |\theta_{i,j} \theta_{j,i}|, \tag{15}$$

where θ^0 is defined as the null matrix of size n if there is no prior information on the GRN. The coefficient in front of the product $|\ \theta_{ji}\theta_{ij}\ |$ is chosen small enough to ensure that if both θ_{ji} and θ_{ij} have been detected at a timepoint, it would generally cost more to put one back close to 0 than to keep it at its computed value.

Third, the temporal dynamics of the process, when available, must be considered. As developed in Section 3.3, the basins that have to be taken into account are all the basins identified during the differentiation process, *i.e* in the different snapshots of the time-varying distribution. But three principal reasons lead to consider that we should not try to infer the network from all the basins at the same time:

1. We can generally see the effect of a GRN in a cell as a signal which is transmitted to the genes [5]. Before the signal has been completely transmitted in the network, many genes are likely to be in a state which does not reflect the GRN. At these moments, the cell is far from the equilibrium. For this reason we would like to take into account the first timepoints, where some genes have not seen the signal, in a weaker way than the last timepoints, where the signal has been well transmitted to all the genes.

2. A population of cells is likely to be more often observed far from the attractors before it has reached its equilibrium than after reaching it. This is due to the fact that when the distribution is far from the equilibrium, some shallow or even unstable basins can be explored, from which the cell can easily escape. In particular, the regions around bifurcations in the cell lineage are expected to be often identified as shallow basins. These basins should then almost be erased if deeper basins are explored afterwards. This also suggests that the first timepoints have to be taken into account in a weaker way than the final ones.
3. Finally, recalling that GRN inference is known to be generally a non-identifiable problem, we simply cannot neglect the information that is brought by temporal information when it is available.

We thereby adopt an iterative approach: for each timepoint, the network is actualized by minimizing the function (14) with the penalization (15), taking as initial condition the network inferred at the previous timepoint. Then, we do not penalize the network itself but its variations to the initial condition. This should satisfy the points 1. and 2., because an erroneous interaction that would have been catched at an early timepoint would be conserved in the final network only if it does not appear in contradiction with the interactions that are inferred afterwards.

Finally, the regression step consists in solving successively, for $k = 1, \cdots, l$, the problem

$$\theta^k = \arg\min_{\theta} R^k(\theta, \alpha_M^k) + \lambda|\theta - \theta^{k-1}|, \qquad (16)$$

where θ^0 is defined as the null matrix of size n, $|\theta - \theta^{k-1}|$ is given by (15), λ is a penalization coefficient and the function R is defined for every θ and $\alpha_M^k = \left\{ \left(\mu_{t_k}(z)\right)_{z\in Z}, (\alpha_z)_{z\in Z} \right\}$ by:

$$R^k(\theta, \alpha_M^k) = \sum_{i=1}^{n} \sum_{z\in Z} \mu_{t_k}(z) \left(\sigma_i^\theta(\frac{\alpha_z}{\alpha_1}) - \frac{\alpha_{z,i}}{\alpha_{1,i}} \right)^2. \qquad (17)$$

The procedure is illustrated in Figure 5. For the applications presented in Section 6, the value of the coefficient λ has been calibrated in order to be optimal for various datasets simulated from randomly-generated tree-like networks, like those studied in Section 6.6.

Interestingly, the cost function (17) can be obtained as the particular case of a more general method for linking a GRN to a set of mixture parameters in the case where single-cell proteomic data were available, and that mRNAs are seen as a proxy for the proteins level. We present this method in Appendix C.

We underline the fact that the function σ^θ, defined in (2), depends only on the GRN θ and $\frac{\alpha_z}{\alpha_1}$, which can be directly estimated from scRNA-seq data, using the method described in Section 5.2.1 and the relations (12). Thus, we do not need to make any assumption on the value of the hyperparameters of the model, not even the ratio $\frac{d_0}{d_1}$ as it was the case in [15]. This is due to the fact that we do not infer the protein distribution but only the values of its main modes, which are completely characterized by the mean of the proteins distribution within each basin, and do not depend on its variance.

The two-steps method presented in Sections 5.2.1 and 5.2.2, that we call CARDAMOM (Cell types Analysis from scRna-seq Data Achieved from a Mixture MOdel), shows good results for the simulated datasets that are presented in Section 6.

5.2.3. Back to the model: consistency of the algorithm and verifications

The inference method presented above corresponds to the calibration of the mechanistic model (1) driven by burst rates functions of the form (2). Recalling that GRN inference is generally not an identifiable problem, many networks are likely to reproduce a dataset. Beyond the topology of the network that we expect to be well reconstructed by the algorithm, a GRN given by CARDAMOM should be considered as accurate if it allows to reproduce the datasets used for the inference, seen as partial observations of the time-varying distribution associated to the mechanistic model. It would then be natural to use the quantitative parameters inferred with CARDAMOM to simulate new snapshot data that we could compare to the original data. However, we point out the fact that the inference method described in Section 5 only used the attractors of the metastable basins associated to the deterministic limit (4), while we recall that the dynamics of the model is mainly determined by the transition rates of the coarse-grained model on the basins, which depends on the value of the potential on the saddle points of the system (4) rather than the attractors. This is in line with the fact that the vector α is supposed to approximate accurately the potential wells of the landscape but not the energetic barrier between them, which remains out of reach. Thus, the

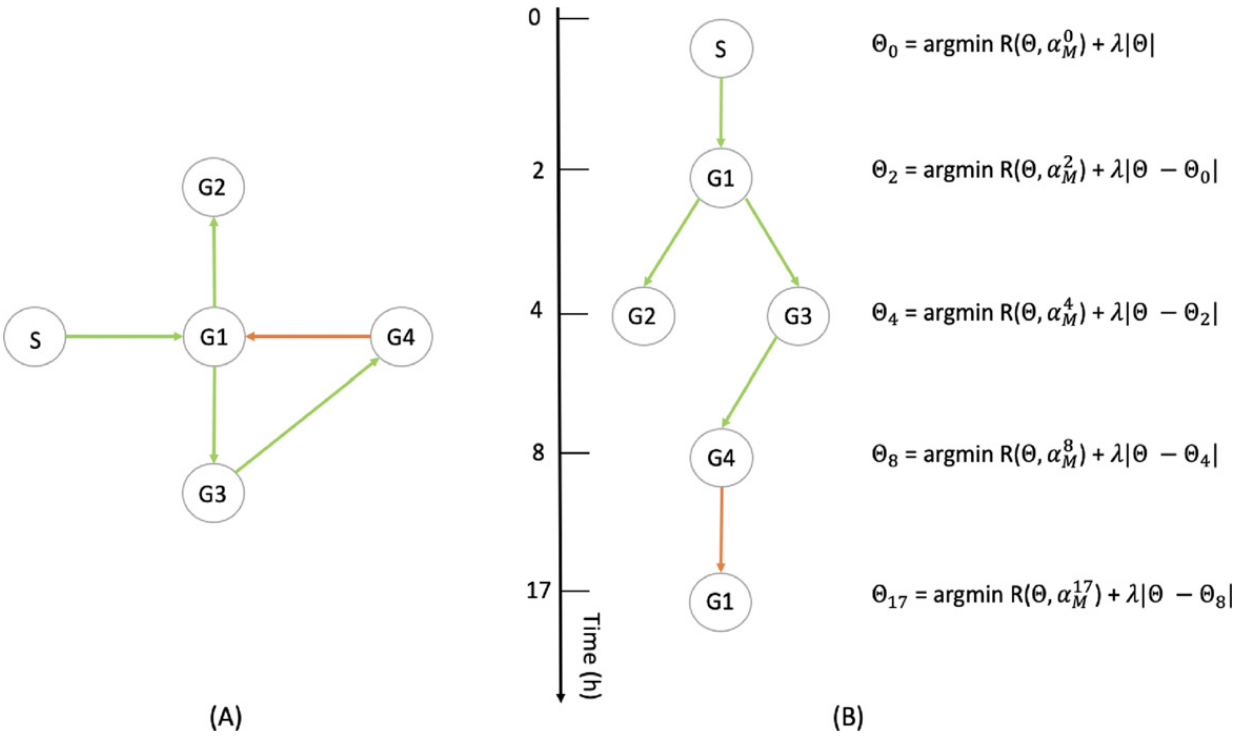


Fig. 5. (A) Example of a 4-genes network (G1 to G4) with a stimulus (S) (see Section 6.1). (B) Illustration of the method described in Section 5.2.2: the interactions being observable only at some particular timepoints, they are progressively inferred, each optimization taking into accounts the interactions that are observed on the previous ones.

method is supposed to find a network which should allow to recover the same metastable basins than the ones observed in the data, the order in which they are visited, but not the right temporal dynamics.

In order to estimate the quality of the inference while taking into account the limitations mentioned above, we should simulate the model (1) for comparing both the initial and the long-time distribution of the model associated to the inferred GRN with the first and the latest snapshot which is available on the data, and verify that simulated cells have a transitory evolution similar to the data (in terms of distribution), even if not necessary with the same temporality. This is the object of Section 6.4.

6. Results

In this section, we present the performance of CARDAMOM on simulated datasets from 3 different types of networks. For each network, the datasets are obtained by sampling independent cells at a certain number of timepoints after applying a stimulus.

6.1. Simulating data with stimulus

The simulations of the model (1) are performed with the python package HARISSA presented in [15], which is based on an efficient thinning method for simulating the model (1). For reproducing *in vitro* differentiation processes, we first simulate every network until reaching a first equilibrium before $t = 0$. At $t = 0$, we introduce a new gene called "stimulus", the proteins level of which is artificially maintained at the value of 1. This value corresponds to to its highest possible attractor: indeed, the proteins scaling (11) implies that no coordinates of the attractors can be greater than 1 in the gene expression space. Proteins being much less noisy than mRNA (*i.e* $\frac{d_{0,i}}{d_{1,i}} \gg 1$), the protein level associated to a gene is expected to be close to the attractor associated to the basin it belongs to, and forcing a gene to be fully expressed in the model (1) with the scaling (11) is equivalent to force its value to 1. Then, the stimulus represents the effect of an artificial perturbation in the differentiation environment inducing cells to evolve towards a new equilibrium. The datasets correspond to the sample of independent cells at a sequence of timepoints.

6.2. Inference of a toggle-switch network from a static expression data sampled under the long time distribution

We first evaluated the ability of CARDAMOM to find the parameters of a 4-genes network consisting

in 2 independent toggle-switch of the same form than the one described in Table 1. To this aim, we simulate 10 dataset by sampling 50 independent cells at 2 timepoints 0 and $20h$. In that case, the stimulus has no effect on the system. Inference is then independently performed for every datasets and the mean of all the significant values obtained for each network is presented in Figure 6, compared to the network used for the simulations. We observe in Figure 6 that the algorithm allows to reconstruct very well the topology of the network, detecting which pairs of genes are in relation together and the sign of the interactions. All other coefficients of the GRN matrix are smaller than 0.3, which can be considered as a residual noise with regards to the principal edges. However, the algorithm computes the values of the coefficients up to a multiplicative factor, which is in line with the limitations we pointed out in Section 5.2.3.

We remark that the algorithm slightly overestimates the diagonal coefficients, which correspond to self-regulation parameters of the genes : the problem of the inference of these parameters being notoriously difficult to infer reliably [40], we will not take into account the diagonal coefficient of the GRN matrix for evaluating the algorithm performances in the following.

6.3. Inference of a 4-genes network with branching and feedback loop from timestamped data

We now evaluate CARDAMOM on the 4-gene network described in Figure 5 and Table 2. Although such a small network may appear very simple, it already has some interesting features (branching, feedback loop with inhibition) and is interesting for inference. To this aim, we simulate 10 datasets by sampling independent cells at 10 time points $t = 0, 2, 4, 6, 8, 11, 13, 15, 17,$ and $20h$, with 50 cells per timepoint (then 500 cells in total for each dataset). Inference is independently performed for every dataset and the results are merged into receiver operating characteristic (ROC) and precision-recall (PR) curves that are shown in Figure 7. CARDAMOM turns out to reconstruct very efficiently the topology of the network. We precise that the sign of the interactions that are inferred is well preserved comparing to the original network.

Table 2
Description of the parameters of the 4-genes network used for the inference in Section 6.3. All others parameters are similar than for the toggle-switch network of Table 1, the same for every genes. The i^{th} line (resp. the i^{th} column) of the matrix θ, correspond to the influence of every genes on the gene i (resp. the influence of the gene i on every genes)

θ_{ij}	$\theta_{\cdot,0}$	$\theta_{\cdot,1}$	$\theta_{\cdot,2}$	$\theta_{\cdot,3}$	$\theta_{\cdot,4}$
$\theta_{0,\cdot}$	0	0	0	0	0
$\theta_{1,\cdot}$	10	0	0	0	-10
$\theta_{2,\cdot}$	0	10	10	0	0
$\theta_{3,\cdot}$	0	10	0	10	0
$\theta_{4,\cdot}$	0	0	0	10	0

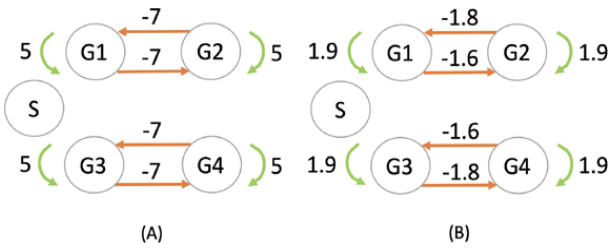


Fig. 6. Inference of a 4-genes network (G1 to G4) consisting in two independent symmetric toggle-switch networks with parameters described in Table 1. The stimulus (S) has no effect on the network, but is nevertheless represented in order to verify whether the inferred network takes it into account or not. The network used for simulating the datasets in (A) is compared with the network inferred by CARDAMOM in (B).

6.4. Reproduction of the data

The bursty model (1) with k_{on} functions defined by (2) can satisfactorily produce expression data for which marginal distributions of genes are close to mixtures of negative binomial distributions, which is known to be biologically relevant. In line with what was discussed in Section 5.2.3, we do not expect the temporal dynamics associated to the network inferred by CARDAMOM to be synchronised with the original data. We then focus on the comparison between the initial and final distributions of the model when simulated by the network presented in Table 2 and the network inferred in Section 6.3, observed at respectively $0h$ and $20h$ after applying the stimulus (see Figure 9). We underline the fact that we do not take the initial distribution of the data to initialize the simulation, but we rather sample a distribution using the inferred network with the stimulus fixed to 0.

We also compare in Figure 10 the two temporal dynamics of the two complete datasets, one used for the inference and one simulated with the inferred network. The data are projected altogether with the method *umap* [27] and we show on each subfigure the cells corresponding to one of the datasets. These cells are then classified depending on the time where the measures are realized after applying the stimulus.

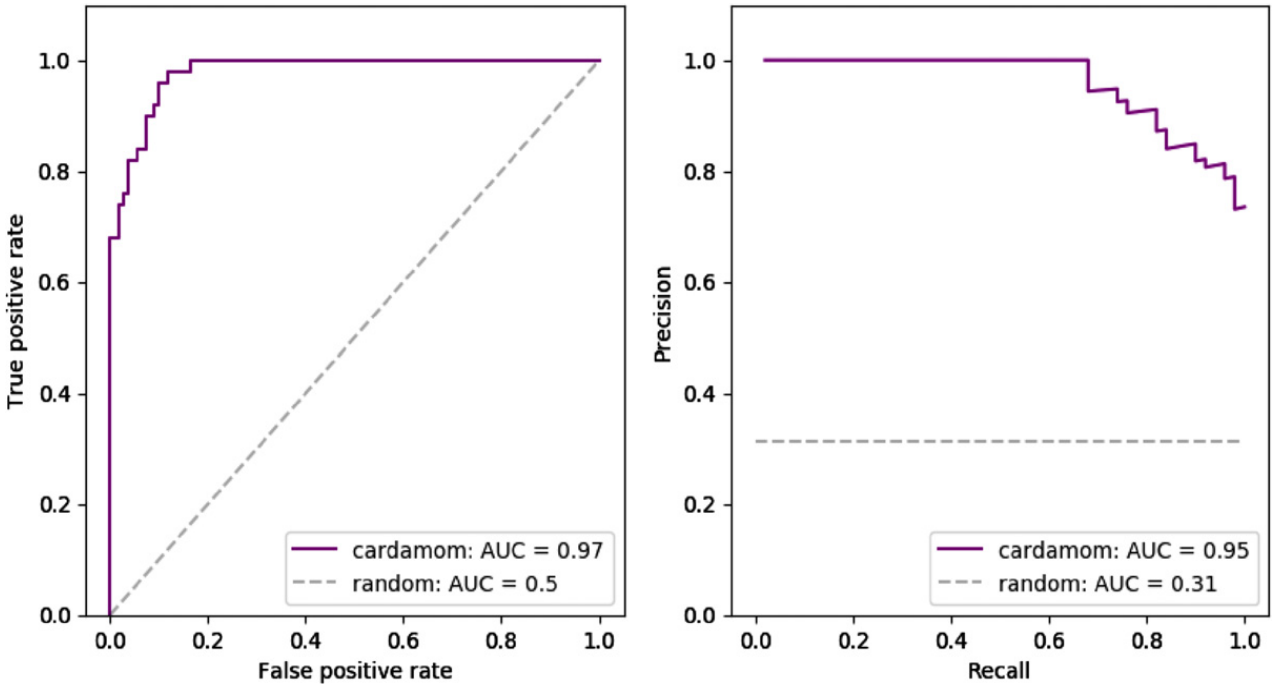


Fig. 7. Inference results for the 4-gene network described in Table 2. Performances are measured in terms of receiver operating characteristic curves (ROC) and precision-recall curves (PR) obtained for 10 independently simulated datasets. Each dataset contained the same 10 timepoints and 50 cells per time point. The dashed gray line indicates the average score that would be obtained by the random estimator (detecting a link or not with equal probability).

The results suggest that the model has been relatively well calibrated through the inference procedure.

6.5. Simplification of the method for sigmoidal burst rate functions

The clustering method described in Section 5.2.1 allows to take into account any number of modes for the burst frequency associated to each gene. However, it is natural to ask if such complexity is compatible with the one allowed by the choice of the functions $k_{on,i}$. In fact, when these functions have the sigmoidal form described in (2), they cannot be expected to be associated to a high number of attractors with intermediate values comprised between $k_{0,i}$ and $k_{1,i}$. A rough simplification consists in considering that every gene is the result of a mixture of the form (6) with only two modes, which can be expected to be close to the values of $k_{0,i}$ and $k_{1,i}$. Following this idea, the clustering step of CARDAMOM can be replaced by a simple binarization of the cells, where the mRNA value measured for each gene is classified as belonging either to the mode $k_{0,i}$ or $k_{1,i}$, with a likelihood ratio between the two negative binomials associated to these parameters. We underline that this approximation does not mean that the data always exhibits strong bimodality. It only means that the fact that they are sometimes observed far from their extremum, which is expected at least during the period of transition after applying the stimulus, does not mean that these transitory states are stable.

The simplification consisting in imposing a number of clusters equal to 2 for each gene would not affect the efficiency of the algorithm for the 4-genes network studied in Section 6.3. Indeed, we observe in Figure 8(A) that the RJMCMC algorithm applied to each gene of the network for every datasets simulated from this network usually computes a number of clusters equal to 2, and in Figure 8(B) that imposing a number of cluster greater than 2 makes slightly decreases the efficiency of the algorithm. We discuss in Section 7.4 the consequences of the accuracy of such simplified method, which may argue in favor of more complex burst functions than the ones of the form (2), or even adaptive functions depending on the clustering step described in Section 5.2.1.

6.6. Application to larger networks

We now consider the cases of tree-like activation networks of 5, 10, 20, 50 and 100 genes. For each case, we simulate ten datasets corresponding to 10 random networks of the same size, that sampled from the uniform distribution over trees rooted in the stimulus [15]. All datasets contain the same 10 timepoints $t = 0, 2, 5, 8, 11, 13, 16, 19, 22,$ and $25h$,

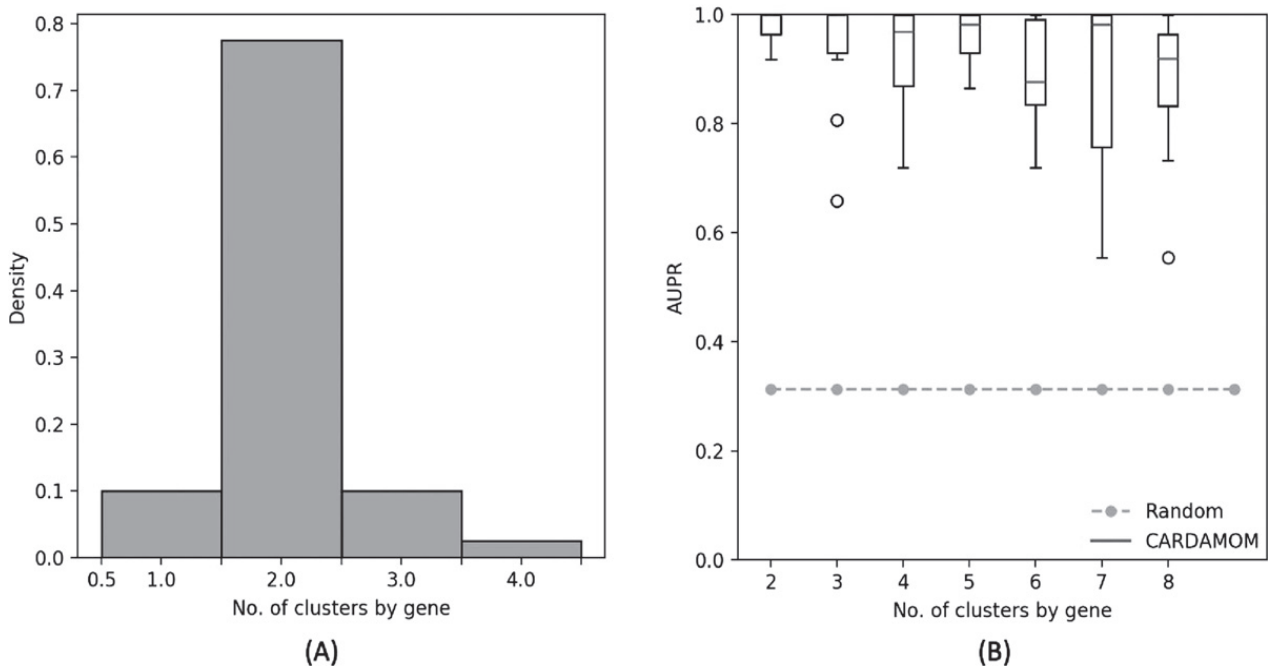


Fig. 8. (A): Distribution of the number of clusters found by the RJMCMC algorithm during the clustering step for each gene of the 4-genes network described in Table 2. (B): Comparison of the performances of CARDAMOM for the same network, when imposing different values for the number of clusters in the clustering step. Performances are measured in terms of area under precision-recall curve (AUPR), based on 10 datasets corresponding to the same network.

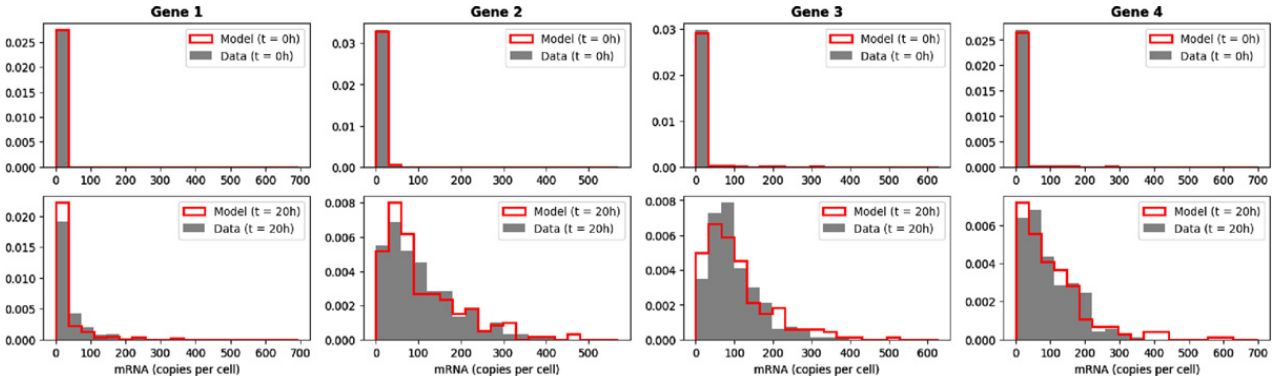


Fig. 9. Comparison of the initial ($t = 0h$) and final ($t = 20h$) empirical marginal distributions of 4 genes simulated by the mechanistic model with the GRN described in Table 2, and the dataset simulated from the GRN inferred by CARDAMOM from the previous dataset (with 200 cells per timepoints).

with 100 cells per timepoint (then 1000 cells in total for each dataset). Inference is then independently performed for all dataset with CARDAMOM and the state-of-the-art method GENIE3 [19]. The results are presented in terms of Area Under Precision-Recall curves (AUPR), in Figure 11A). CARDAMOM turns out to reconstruct more efficiently the topology of the network than GENIE3. The strong decrease in performance when the number of genes gets higher is due not only to the lack of data per timepoints in regards to the number of interactions, but also to the choice of the timepoints. Indeed, a sequence of timepoints that is too coarse to catch the dynamics would lead to a lack of accuracy in the inference, and a sequence which is too tight would often be too short, leading to miss the activity of some genes. This appears to be the case from 20 genes in Figure 11A).

For every size of network, an average runtime is obtained after inferring the 10 datasets associated to the 10 tree-like networks, on a 16-GB RAM, 2,4 GHz Intel Core i5 computer. We see in Figure 11B) that for both algorithms the computational speed increases linearly with respect to the number of genes, with a slope which is significantly lower for CARDAMOM. These results show that the algorithm is suitable for realistic number of genes and cells.

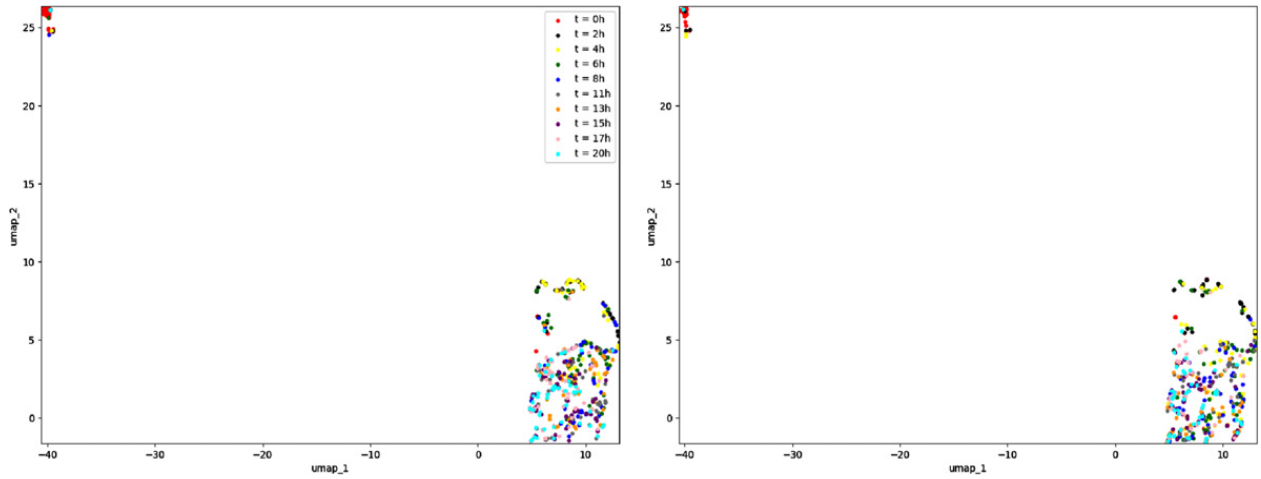


Fig. 10. Representation in 2 dimensions of (A) the dataset used for the inference, simulated from the GRN described in Table 2 and (B) the dataset simulated from the GRN inferred by CARDAMOM. Each dataset contains 500 cells divided in 10 timepoints.

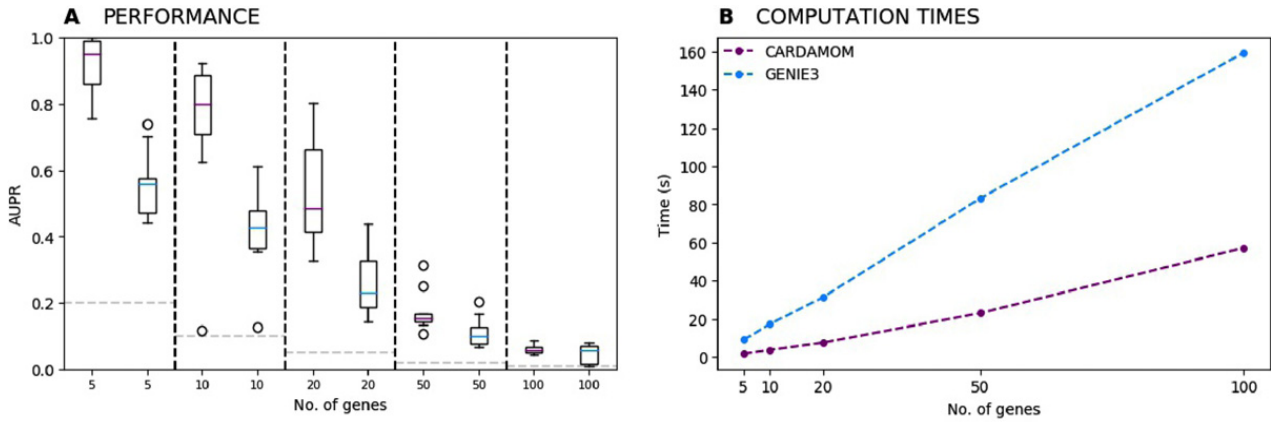


Fig. 11. (A) Evolution of the performances of CARDAMOM and GENIE3 when increasing the number of genes. For each number of genes, we represent a boxplot of the AUPR scores computed for ten datasets simulated with 10 different randomly generated tree-like networks. (B) Evolution of the average computational time measured for inferring the tree-like networks with respect to the number of genes.

7. Discussion and prospects

In this discussion, after clarifying in Section 7.1 the importance of the clustering step of CARDAMOM, we go deeper in Section 7.2 into the link between the mixture approximation and the popular notion of Waddington landscape. Then, we discuss in Section 7.3 the applicability of our method to real single-cell datasets. Finally, we show in Section 7.4 that the regression step of the method highlight the link existing between the mechanistic model and a neural network, and in what way this analogy paves the way for a more general method using adaptive bursts rates functions.

7.1. Importance of the clustering

The attentive reader should have remarked that the function (17) which is optimized in the second step of CARDAMOM appears very simple and close to a generalized linear model, which is surprising for an analysis based on a mechanistic model. It is then natural to ask whether the clustering step remains important or if it is possible to reduce the method to the regression step by identifying each cell to a distinct attractor without losing an important accuracy (the clustering step would then simply consists in rescaling each count of a gene i by the parameters c_i). We compare in Figure 12 this coarse simplification with CARDAMOM on the datasets used in Section 6.3. We observe that such simplification generates a significant decrease in accuracy, confirming that the method cannot be reduced to such generalized regression.

7.2. Limits of the method for finding the developmental landscape

In this section, we come back on the Gamma mixture approximation (6) and its limitations for

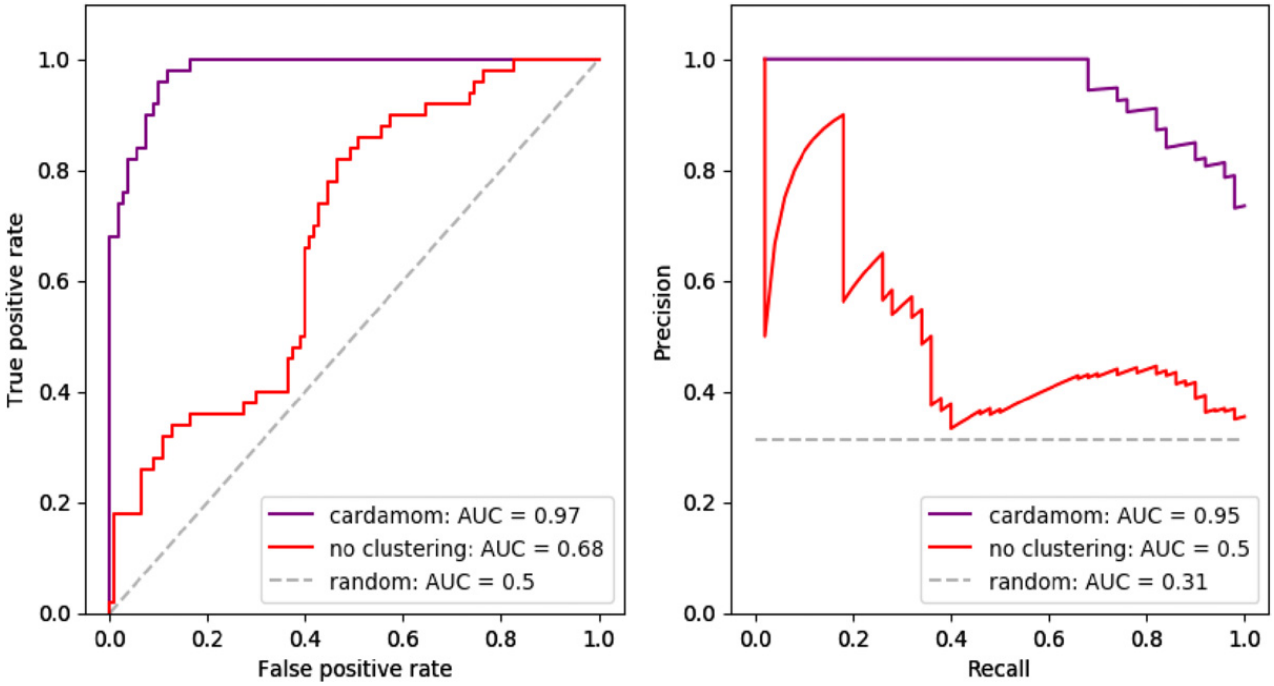


Fig. 12. Analogy of Figure 7, where we compare (A) the ROC curves and (B) the PR curves of the GRN reconstruction performed by CARDAMOM with clustering (in purple) and without clustering (in red, considering that there are as many clusters as cells in each dataset).

describing the developmental landscape of differentiation, in order to remove some of the confusion that might arise from misuse of this approximation. As we presented in Section 3, this mixture approximation is based on two assumptions:

1. The functions $k_{on,i}$ are close to constant inside the basins:
2. The jump of a cells from one basin of attraction of the deterministic limit (4) to another is a rare event.

Then, in terms of Waddington landscape, the mixture is supposed to describe accurately the bottom of the potential wells associated to a GRN and, at least under the assumption 1., the curvature of the potential wells that are deep enough. However, it does not describe accurately the boundaries of these wells, which characterize the energetic barrier and correspond precisely to areas of the gene expression space where the functions $k_{on,i}$ have an important gradient, making the approximation not accurate. In other words, the potential

$$V^{\alpha} = -\ln\left(\sum_{z\in Z}\mu(z)\prod_{j=1}^{n}\mathrm{Gamma}\left(\frac{k_{z,j}}{d_{1,j}}, c_j\right)\right)$$

is not supposed to be a good approximation of the real potential $V^{\theta} = -\ln\left(\hat{u}^{\theta}\right)$ associated to a GRN when a cell is far from the attractors. This explains, with a statistical physic point of view, why our algorithm is not able to catch the temporal dynamics of the data, but only the same sequence of temporal distributions.

This limitation is very hard to overcome. For this sake, we should either get knowledge on the values of the cells distribution on saddle points of the landscape which are located in areas of the gene expression space where it is very unlikely to find cells, which would require to have thousands of cells just for a small network, or to take into account explicitly the value of time in the inference method. The latter point would require to have analytical results on the temporal distribution of cells in the gene expression space, which is generally out of range for realistic models like (1). For example, using the temporal dynamics of the probability vector μ_t on the basins for reconstructing the transition rates (see Section 3.3) would be challenging as an analytical link between these transition rates and the GRN can only be obtained up to a prefactor, which does not depend on the scaling factor ε but is specific to each transition [51]. The only solution seems then to combine analytical results with simulations, in the same philosophy than [5], but this leads to much more complex and time-consuming algorithms that the one which is developed in this article. To our point of view, this also argues against methods where a mixture model is used for reconstructing the temporal dynamics of a metastable process from stationary distributions [37].

7.3. Applicability of CARDAMOM to real single-cell datasets

The algorithm CARDAMOM has been shown to reconstruct accurately a most-likely GRN from timestamped *in silico* datasets, through the estimation of the metastable parameters associated to these data. It is then natural to ask whether the method could deal with real datasets or not. The method for estimating the metastable parameters should be well suitable for real datasets, provided that the Negative Binomial mixture approximation of mRNAs is accurate, and we have seen in Section 6.6 that the algorithm is suitable for realistic number of genes and cells. The main difficulty that could appear would then be related to the decrease in the performance of the algorithm with the number of modes appearing in the sample. Indeed, we recall that the regression step aims to infer a GRN matrix of size $n \times n$ from the values of the functions $k_{on,i}$ at each attractor. First, it implies that too few clusters in the sample would mean too little information for the inference. Second, as developed in Section 6.5, too many clusters for each gene would make the sigmoidal bursts rate functions unsuitable, which could also lower the performances of the algorithm. Then, the level of multistability seems then to be critical for the accuracy of the inference. For the first point, we argue that regarding the noisy nature of mRNA counts, it is impossible to infer a reliable GRN when there is not enough multistability in the sample, *i.e* when the marginal distribution of mRNA associated to each gene is described by a simple negative binomial, the parameters of which does not evolve in time. We discuss in Section 7.4 the implications of the second point in terms of modeling.

To go further, we recall that the mathematical analysis beyond CARDAMOM lays on the point of view that cell differentiation processes can be coarsely reduced into a discrete process on a limited number of cell types, *i.e* that the main ingredients for characterizing stochasticity are the frequency modes describing the cell types and the random transitions between them, which is commonly accepted since [18]. Such transitions have been for example recently proposed as the basis for facilitating the concomitant maintenance of transcriptional plasticity and stem cell robustness [55]: in this case, the authors had proposed a phenomenological view of the transition dynamics between states, and our work may typically connect this cellular plasticity to an underlying GRN dynamics. This connection between a GRN and associated cell states should also be used for the quantitative modeling of stochastic state transitions underlying the generation of diversity in cancer cells [14, 56], or for investigating transitions between the large diversity of clusters that can be observed in human ES cell differentiation datasets [29].

7.4. Gene expression as a neural network

Interestingly, CARDAMOM appears close to a machine learning approach. Indeed, since each function $k_{on,i}$ is independent from the others, depending only on the i^{th} line of the matrix θ, the regression step is fairly close to parallel regressions of each gene on the others. The choice of sigmoidal bursts rate functions makes these regressions appear as the learning step of an artificial neural network, which is simply a set of one-layer perceptron for each gene coupled by the crossed-penalization between the symmetric coefficients of the matrix θ (see (15)). In the light of this framework, the first step of the algorithm can be seen as an identification of the outputs from the data, by identifying the mode associated to each mRNA count. For sigmoidal bursts rate functions, we have seen in Section 6.5 that it is reasonable to consider only two modes. This makes the clustering step close to the prepossessing of a Boolean network analysis, which consists in assigning every cell to the state of a Boolean model, where a 0 may rather take any values between 0 and 1. Without crossed-penalization, the second step would exactly consist in learning the parameters of a perceptron as represented in Figure 13, where each line of the matrix θ corresponds to the weight of the perceptron. The method then builds an interesting link between machine learning and dynamical modeling.

The results presented in Section 6.5 let us suppose that the sigmoidal bursts rate functions of the form

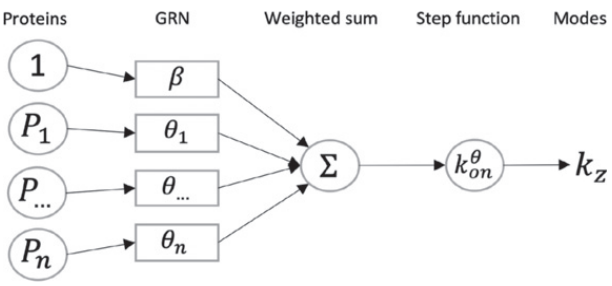


Fig. 13. CARDAMOM can be interpreted as the learning of a perceptron, where the clustering step (Section 5.2.1) corresponds to the classification of the data, regarding to their associated modes of frequency, and the regression step (Section 5.2.2) to the identification of the weights of the perceptron, corresponding to the GRN.

(2) have a limited complexity, *i.e* that they are not able to catch a multistability which would be associated to more than two potential wells for each gene. Thus, when applying the clustering step to a dataset showing an important multimodality, at least for some genes, the sigmoidal function may be not adapted for modeling the behaviour of the underlying process.The clustering step of CARDAMOM, described in Section 5.2.1 could therefore be seen as a preliminary to the choice of the "right" bursts rate function that should be used for minimizing the risk (17) (see Section 5.2.2): this function should be chosen in accordance with the number of modes detected for each gene. The neural network framework introduced above is interesting, as it suggests that the choice of the bursts rate function could correspond to the choice of the number of layers in the neural network of Figure 13. In that point of view, the work achieved in this paper treats the basic case of one layer, which could be extended, although the interpretability of the nodes for more than one layer remains an open question. We believe that this adaptability is necessary when dealing with highly complex single-cell data.

8. Conclusion

We proposed in this work an efficient method for performing GRN inference from single-cell data, seen as the calibration of a mechanistic model of gene expression involving mRNA and protein levels. To the best of our knowledge, our approach is the first one which uses explicitly the popular notion of approximate developmental landscape, through the concept of metastability, for performing the inference. The method relies on a previous analysis of the mechanistic model developed in [51]. It provides a modular algorithm which consists in two steps: in a first time characterizing the metastable parameters associated to the coarse-grained model which describes the best the data, and in a second time solving a serie of regression problems aiming to link these parameters to a most-likely GRN. The method is implemented as a Python package called CARDAMOM, which is available on open-access. This algorithm seems to be very accurate for small but complex networks, and its computational speed allows to make it suitable for realistic number of genes and cells. Furthermore, such inference method based on a mechanistic model has the great advantage to make quantitative predictions that can then be easily tested by simulating the model. The simplicity of the regression step is particularly remarkable, and appears close to the learning step of a neural network. We believe that in addition to efficiently infer GRNs from timestamped datasets while keeping a high interpretability, CARDAMOM, which combines explicitly machine learning methods with mathematical modeling, could pave the way for new adaptive methods.

The next step would be to conceive bigger and more complex networks for which all the non-zeros entries of the GRN matrix do have an impact on the data when simulating the model, which is in itself a challenge, and to compare the performances of CARDAMOM to state-of-the-art algorithms used for GRN inference. As the problem of inference from a mechanistic model like (1) has been recently studied elsewhere using distinct mathematical tools [15], a collaboration is in progress for realizing an important benchmark of our method together with another algorithm specifically developed for facing transcriptional bursting and well-known methods such as GENIE 3 [19] and PIDC [8], and studying the ability of the mechanistic model to reproduce *in vitro* expression datasets when the model is calibrated by CARDAMOM.

Code availability

The algorithm CARDAMOM presented in Section 5.2, as well as the code for generating the Figure 7, are available on https://gitbio.ens-lyon.fr/eventr01/cardamom.

The algorithm used for simulating the model as well as to generate the tree-like networks used in Section 6.6 can be found on https://github.com/ulysseherbach/harissa.

Acknowledgment

This work was supported by funding from French agency ANR (SingleStatOmics; ANR-18-CE45-0023-03). I would like to thank especially Thibault Espinasse for enlightening discussion and statistical advices, as well as Ulysse Herbach for having highlighted the notion of bursts modes for the model of gene expression, Thomas Lepoutre and Olivier Gandrillon for critical reading of the manuscript. I also thank all members of the SBDM and Dracula teams, and of the SingleStatOmics project, for providing such stimulating working environment. I finally thank the BioSyL Federation, the LabEx Ecofect (ANR-11-LABX-0048) and the LabEx Milyon of the University of Lyon for inspiring scientific events.

References

[1] K. Akers and T.M. Murali, Gene regulatory network inference in single cell biology, In: *Current Opinion in Systems Biology* (2021).

[2] C. Albayrak, et al., Digital quantification of proteins and mRNA in single mammalian cells. In: *Molecular Cell* **61**(6) (2016), pp. 914–924.

[3] V. Antolovic, et al., Generation of Single-Cell Transcript Variability by Repression, In: *Curr Biol* **27**(12) (2017), 1811–1817 e3.

[4] M. Bizzarri, et al., Gravity Constraints Drive Biological Systems Toward Specific Organization Patterns: Commitment of cell specification is constrained by physical cues. In: *Bioessays* (2017).

[5] A. Bonnaffoux, et al., WASABI: a dynamic iterative framework for gene regulatory network inference, In: *BMC Bioinformatics* **20**(1) (2019), pp. 1–19.

[6] N. Bouguila and T. Elguebaly, A fully bayesian model based on reversible jump MCMC and finite beta mixtures for clustering, In: *Expert Systems with Applications* **39**(5) (2012), pp. 5946–5959.

[7] E. Braun, The unforeseen challenge: from genotype-to-phenotype in cell populations. In: *Rep Prog Phys* **78**(3) (2015), pp. 036602.

[8] T.E. Chan, M. Stumpf and A.C. Babtie, Gene regulatory network inference from singlecell data using multivariate information measures, In: *Cell Systems* **5**(3) (2017), pp. 251–267.

[9] A.F. Coskun, U. Eser and S. Islam, Cellular identity at the single-cell level. In: *Mol Biosyst* **12**(10) (2016), pp. 2965–2979.

[10] J.G. Dias and M. Wedel, An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods, In: *Statistics and Computing* **14**(4) (2004), pp. 323–332.

[11] A. Faggionato, D. Gabrielli and M. Crivellari, Non-equilibrium thermodynamics of piecewise deterministic Markov processes. In: *Journal of Statistical Physics* **137**(2) (2009), pp. 259.

[12] N.P. Gao, et al., Universality of cell differentiation trajectories revealed by a reconstruction of transcriptional uncertainty landscapes from single-cell transcriptomic data, In: *bioRxiv* (2020).

[13] A. Guillemin, et al., Drugs modulating stochastic gene expression affect the erythroid differentiation process, In: *PLOS ONE* **14**(11) (2019), e0225166.

[14] P.B. Gupta, et al., Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells, In: *Cell* **146**(4) (2011), pp. 633–44.

[15] U. Herbach, Gene regulatory network inference from single-cell data using a self-consistent proteomic field. In: *arxiv* (2021).

[16] U. Herbach, et al., Inferring gene regulatory networks from single-cell data: a mechanistic approach, In: *BMC Systems Biology* **11**(1) (2017), pp. 105. issn: 1752-0509.

[17] S. Huang and D. Ingber, A non-genetic basis for cancer progression and metastasis: selforganizing attractors in cell regulatory networks, In: *Breast Disease* **26**(1) (2007), pp. 27–54.

[18] S. Huang, et al., Cell fates as high-dimensional attractor states of a complex gene regulatory network, In: *Physical Review Letters* **94**(12) (2005), pp. 128701.

[19] A. Irrthum, L. Wehenkel, P. Geurts, et al., Inferring regulatory networks from expression data using tree-based methods, In: *PloS One* **5**(9) (2010), e12776.

[20] S. Kauffman, A proposal for using the ensemble approach to understand genetic regulatory networks, In: *Journal of Theoretical Biology* **230**(4) (2004), pp. 581–590.

[21] Minoru SH Ko, A stochastic model for gene induction, In: *Journal of Theoretical Biology* **153**(2) (1991), pp. 181–194.

[22] G.-W. Li and X.S. Xie, Central dogma at the single-molecule level in living cells. In: *Nature* **475**(7356) (2011), pp. 308–315.

[23] Y.T. Lin and T. Galla, Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models, In: *Journal of The Royal Society Interface* **13**(114) (2016), pp. 20150772.

[24] M. Mackey, M. Tyran-Kamińska and R. Yvinec, Molecular distributions in gene regulatory dynamics, In: *Journal of Theoretical Biology* **274**(1) (2011), pp. 84–96.

[25] J. Mar, The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond, In: *Biophys Rev* (2019), pp. 89–94.

[26] J. Mar, The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond, In: *Biophysical Reviews* **11**(1) (2019), pp. 89–94.

[27] L. McInnes, J. Healy and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, In: *arXiv preprint arXiv:1802.03426* (2018).

[28] M. Mojtahedi, et al., Cell Fate Decision as High-Dimensional Critical State Transition, In: *PLoS Biol* **14**(12) (2016), e2000640.

[29] K. Moon, et al., Visualizing structure and transitions in high-dimensional biological data, In: *Nature Biotechnology* **37**(12) (2019), pp. 1482–1492.

[30] N. Moris and A.M. Arias, The Hidden Memory of Differentiating Cells, In: *Cell Syst* **5**(3) (2017), pp. 163–164.

[31] N. Moris, C. Pina and A.M. Arias, Transition states and cell fate decisions in epigenetic landscapes, In: *Nat Rev Genet* **17**(11) (2016), pp. 693–703.

[32] S.A. Morris, The evolving concept of cell identity in the single cell era. In: *Development* **146**(12) (2019), pp. 1–5.

[33] A. Moussy, et al., Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment, In: *PloS Biol* **15**(7) (2017), e2001867.

[34] D. Nicolas, N.E. Phillips and F. Naef, What shapes eukaryotic transcriptional bursting? In: *Mol Biosyst* **13**(7) (2017), pp. 1280–1290.

[35] H. Ochiai, et al., Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells, In: *Scientific Reports* **4**(1) (2014), pp. 1–9.

[36] G. Papanicolaou, Asymptotic analysis of transport processes. In: *Bulletin of the American Mathematical Society* **81**(2) (1975), pp. 330–392.

[37] P. Pearce, et al., Learning dynamical information from static protein and sequencing data, In: *Nature Communications* **10**(1) (2019), pp. 1–8.

[38] J. Peccoud and B. Ycart, Markovian modeling of gene-product synthesis. In: *Theoretical Population Biology* **48**(2) (1995), pp. 222–234.

[39] A. Pratapa, et al., Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data, In: *Nat Methods* **17**(2) (2020), pp. 147–154.

[40] A. Pratapa, et al., Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data, In: *Nature Methods* **17**(2) (2020), pp. 147–154.

[41] A. Richard, et al., Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process, In: *PloS Biol* **14**(12) (2016), e1002585.

[42] S. Richardson and P.J. Green, On Bayesian analysis of mixtures with an unknown number of components (with discussion), In: *Journal of the Royal Statistical Society: series B (statistical methodology)* **59**(4) (1997), pp. 731–792.

[43] J. Rodriguez and D.R. Larson, Transcription in Living Cells: Molecular Mechanisms of Bursting, In: *Annu Rev Biochem* **89** (2020), pp. 189–212.

[44] G. Sanguinetti, et al., Gene regulatory network inference: an introductory survey, In: *Gene Regulatory Networks. Springer*, (2019), pp. 1–23.

[45] Abhishek K. Sarkar and M. Stephens, Separating measurement and expression models clarifies confusion in single cell RNA-seq analysis, In: *BioRxiv* (2020).

[46] S. Semrau, et al., Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells, In: *Nat Commun* **8**(1) (2017), pp. 1–16.

[47] P.S. Stumpf, et al., Stem Cell Differentiation as a Non-Markov Stochastic Process, In: *Cell Systems* **5** (2017), pp. 268–282.

[48] D.M. Suter, et al., Mammalian genes are transcribed with widely different bursting kinetics, In: *Science* **332**(6028) (2011), pp. 472–474.

[49] A. Teschendorff and A. Feinberg, Statistical mechanics meets single-cell biology, In: *Nature Reviews Genetics* (2021), pp. 1–18.

[50] E. Tunnacliffe and J. Chubb, What Is a Transcriptional Burst? In: *Trends Genet* **36**(4) (2020), pp. 288–297.

[51] E. Ventre, et al., Reduction of a stochastic model of gene expression: Lagrangian dynamics gives acces to basins of attraction as cell types and metastability. In: *bioRxiv* (2020).

[52] C.-H. Waddington, *The Strategy of the Genes*, Routledge, 1957.

[53] J. Wang, et al., Quantifying the Waddington landscape and biological paths for development and differentiation, In: *Proceedings of the National Academy of Sciences* **108**(20) (2011), pp. 8257–8262.

[54] J. Wang, et al., The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation, In: *Biophysical Journal* **99**(1) (2010), pp. 29–39.

[55] J.C. Wheat, et al., Single-molecule imaging of transcription dynamics in somatic stem cells, In: *Nature* (2020).

[56] J.X. Zhou, et al., Nonequilibrium population dynamics of phenotype conversion of cancer cells, In: *PLoS One* **9**(12) (2014), e110714.

[57] J.X. Zhou, et al., Quasi-potential landscape in complex multi-stable systems, In: *Journal of the Royal Society Interface* **9**(77) (2012), pp. 3539–3553.

[58] C. Zong, et al., Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene, In: *Molecular Systems Biology* **6**(1) (2010), pp. 440.

Appendix A Description of the parameters used for the model

We provide here a list of the main variables that are used throughout the article.
1. For the gene expression model:

- θ a matrix defining the interactions between genes, corresponding to a matrix with diagonal terms defining external stimuli,
- $k_{0,i}$ is the basal rate of expression of gene i,
- $k_{1,i}$ is the maximal rate of expression of gene i,
- β_i is the basal activity of gene i, which can be also considered as the constant activity of set of genes which are not measured and act on the network,
- $d_{0,i}$ is the degradation rates for mRNAs of gene i,
- $d_{1,i}$ is the degradation rate for proteins of gene i,
- $s_{0,i}$ is the creation rate for mRNAs of gene i,
- $s_{1,i}$ is the creation rate for proteins of gene i,
- $k_{off,i}$ is the exponential rate of switching from state *on* to state *off* for the promoter of gene i,
- $k_{on,i}$ is a sigmoidal function depending on the global protein field $P \in \Omega$, defined in (2), which characterizes the exponential rate of switching from state *off* to state *on* for the promoter of gene i,
- $c_i = \frac{k_{off,i} d_{0,i}}{s_{0,i} s_{1,i}}$ is the exponential rate of proteins of gene i that are created at every burst, in the model (3).

2. For the numerical analysis allowing to understand the algorithm CARDAMOM:

- Z denotes a set of basins of attraction associated to the deterministic limit for a given GRN, and $(P_z)_{z\in Z}$ the set of corresponding attractors,
- $k_{z,i} = k_{on,i}(P_z)$ corresponds to frequency mode of the promoter of gene i associated to the basin $z \in Z$,
- μ_t is a probability vector describing the probability for a cell to be in each basin $z \in Z$ at time t. μ denotes this probability vector at the steady state,
- $\alpha = \left\{\mu(z), \frac{k_{z,i}}{d_{1,i}}, c_i\right\}_{z\in Z, i=1,\cdots,n}$ is the set of parameters which describe a Gamma mixture of the form (6),
- α_M^k is the set of parameters which describe a Gamma mixture associated to a gene expression matrix M_{t_k} measured at time t_k,
- $\alpha_{z,i} = \frac{k_{z,i}}{d_{0,i}}$ denotes the renormalized frequency mode for the promoter of gene i of a cell within a basin $z \in Z$, that is accessible from scRNA-seq data.

Appendix B Master equation of the reduced model

The master equation on the probability density $u(t, \cdot)$ of the bursty model (3), describing only proteins, associated to a GRN θ appears as an integro-differential equation:

$$\partial_t u(t,x) = \sum_{i=1}^{n} \left[\partial_{x_i} \left[d_{1,i} x_i u(t,x) \right] + \int_0^{x_i} k_{on,i}^{\theta}(x - he_i) u(t, x - he_i) c_i e^{-c_i h} dh - k_{on,i}^{\theta}(x) u(t,x) \right], \tag{18}$$

where each e_i is a vector of size n with only zero entries except on the i^{th} position.

Appendix C Regression method in case of proteomic data

In this section, we show how CARDAMOM, which aims to infer a GRN from scRNA-seq data, can be interpreted as a restriction of a more general algorithm where mRNAs are seen as a proxy for the proteins levels, which uses more intensively the characteristics of the Gamma mixture approximation (6) instead of the modes only.

In the model (1), a GRN is completely encoded through the functions $k^{\theta}_{on,i}$. Then, it is straightforward that knowing the exact values taken by these functions on the whole protein space would allow to determine the value of the associated GRN. The method described in Section 4.1 consists in using the value of these functions on the set of the attractors, but we did not directly use the information provided by the probability vector on the basins μ. Given a experimentally observed distribution of proteins, it would be fruitful if the mixture approximation allowed to evaluate the functions $k_{on,i}$ not only on the attractors of the basins but also on the position of the cells that are observed. For this sake, we define the functions $k^{\alpha}_{on,i}$ for all $i = 1, \cdots, n$, for all $x \in \Omega$:

$$k^{\alpha}_{on,i}(x) = \frac{\sum_{z \in Z} \mu_z k_{z,i} \prod_{j=1}^{n} \text{Gamma}(k_{z_j}, c_j)(x)}{\sum_{z \in Z} \mu_z \prod_{j=1}^{n} \text{Gamma}(k_{z_j}, c_j)(x)}. \tag{19}$$

The two following lemmas are proved at the end of this section:

Lemma 1. Replacing $k^{\theta}_{on,i}$ by $k^{\alpha}_{on,i}$ in the model (3), the stationary distribution is exactly given by the mixture distribution (6).

Lemma 2. We consider the evolution of a population of cells, initially distributed under a Gamma mixture $\hat{u}_{\alpha}$ of the form (6), in the model (3) driven by the functions $k^{\theta}_{on,i}$. We have the inequality:

$$\forall x \in \Omega, \int_{\Omega} |\partial_t \hat{u}_{\alpha}(x)|_{|t=0} dx \leq 2\mathbb{E}_{\hat{u}_{\alpha}} ||k^{\theta}_{on}(X) - k^{\alpha}_{on}(X)||_1.$$

Lemma 1 suggests that the difference between the distributions $\hat{u}_{\theta}$ and $\hat{u}_{\alpha}$, which is supposed to be small according the the analysis of Section 3, should be related to the difference between the two classes of parametric functions $k^{\theta}_{on,i}$ and $k^{\alpha}_{on,i}$. The inequality of Lemma 2 gives a more precise reason for considering this difference. Indeed, the expected value of the 1-norm of the difference between these two functions is an upper bound of a quantity which measures how rapidly the associated mixture distribution is going to change when it is taken as an initial condition in the master equation of the model (3) driven by $k^{\theta}_{on,i}$, that we call the impulsion of this mixture. We remark that for every product of Gamma distributions centered on one of the attractors associated to a GRN, the expected value on the right hand side of the inequality of Lemma 2 is expected to be small, as the cells are not going to jump to another basin in short times. The "right" mixture associated to a GRN θ is then the one which is going to be accurate in the highest part of the gene expression space, *i.e* that takes into account the highest number of basins. But a simple sum of Gamma mixture would not be accurate on the areas of the gene expression space where the Gamma distributions cross each other, as it does not take into account the potential depth associated to each attractor. The balance $\mu(z)$ can then be seen as a balance which characterizes the most-likely energetic barrier between the potential wells for reconstructing the right steady-state behavior. Note that this ability of the vector μ to identify the depth of the basins should be closely linked to the ability of a quasipotential to describe the transitions between the basins [51].

We compare in Figure 14 the functions $k^{\theta}_{on,1}$ and $k^{\alpha}_{on,1}$ for the toggle-switch network described in Table 1, illustrating the fact that the approximation seems accurate inside the basins and allows to reproduce quite accurately the basins of attraction of the deterministic system, although the functions on the boundary of the basins have a slightly different behaviour. This is in line with the ideas developed in Section 5.2.3.

The upper bound of the impulsion of a mixture distribution in the model driven by a GRN θ, which is provided by the inequality of Lemma (2), appears as a good choice for the function R described in Section 4.2. It is natural to substitute the distribution $\hat{u}_{\alpha(X)}$ by the empirical distribution associated to the matrix X. We also decide to take the 2-norm in order to simplify the minimization problem. We would finally obtain instead of (17):

$$R(\theta, \alpha(X)) = \sum_{x \in X} \sum_{i=1}^{n} \left(k^{\theta}_{on,i}(x) - k^{\alpha(X)}_{on,i}(x) \right)^2. \tag{20}$$

We remark that the function to be optimized in CARDAMOM (17) corresponds exactly to this new function (20) when the cells x are exactly located on the attractors. This is in line with the fact that in our framework, the

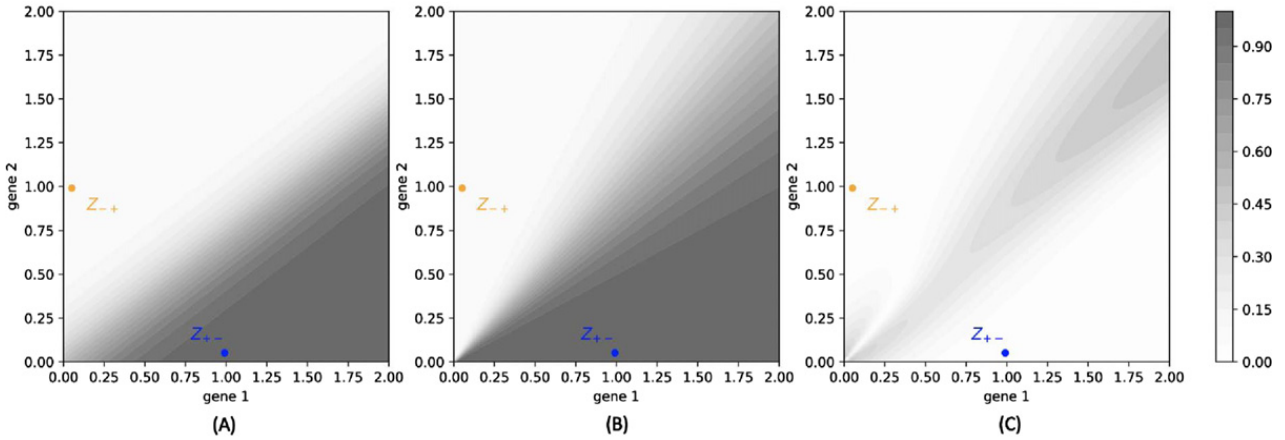


Fig. 14. (A): Color map of the function $k^{\theta}_{on,1}/k_{1,1}$ characterized by the toggle-switch described in Table 1, on the gene expression space. (B): Color map of the function $k^{\alpha}_{on,1}/k_{1,1}$ characterized by the mixture parameters associated to the same network, obtained with the method described in Figure 4, on the gene expression space. (C): Color map of the function $|k^{\theta}_{on,1} - k^{\alpha}_{on,1}|/k_{1,1}$, on the gene expression space.

most likely position for the proteins vector P knowing mRNAs are on the attractors of the basins to which the cell belongs. This lack of information cannot be compensated: indeed, it would either suppose that an observed deviation of a count of mRNA from an attractor is better explained by the fact that the proteins are far from their most-likely position than mRNA itself, which is not the case since mRNAs are known to be much more noisy than proteins, either that the functions $k_{on,i}$ are far from their value on the attractors even when proteins are far from the boundary, in which case the Gamma mixture approximation (6) would not be accurate and our method is no more relevant.

Proof of Lemma 1

Injecting the functions $k^{\alpha}_{on,i}$ instead of $k^{\theta}_{on,i}$ and the distribution $\hat{u}^{\alpha}$ of the form (6) instead of u on the right-hand side of the master equation (18), we obtain for all $i = 1, \cdots, n$:

$$\partial_{x_i}\left[d_{1,i}x_i\hat{u}^{\alpha}(x)\right] + \int_0^{x_i} k^{\alpha}_{on,i}(x - he_i)\hat{u}^{\alpha}(x - he_i)c_i e^{-c_i h}dh - k^{\alpha}_{on,i}(x)u(t,x) \quad =$$

$$\sum_{z\in Z}\mu(z)\prod_{j\neq i}\text{Gamma}\left(\frac{k_{z,j}}{d_{1,j}}, c_j\right)\left[\partial_{x_i}\left[d_{1,i}x_i\text{Gamma}\left(\frac{k_{z,i}}{d_{1,i}}, c_i\right)(x)\right] + \right.$$

$$\left.\int_0^{x_i} k_{z,i}\text{Gamma}\left(\frac{k_{z,i}}{d_{1,i}}, c_i\right)(x - he_i)c_i e^{-c_i h}dh - k_{z,i}\text{Gamma}\left(\frac{k_{z,i}}{d_{1,i}}, c_i\right)(x)\right] = 0,$$

because $\text{Gamma}\left(\frac{k_{z,i}}{d_{1,i}}, c_i\right)$ is the unique stationary distribution of the model (3) for one gene and a constant $k_{on,i} = k_{z,i}$ [24]. Then, the right-hand side of the master equation is null and $\hat{u}^{\alpha}$ is the unique stationary distribution of the model (3) when the burst rates functions are of the form (19).

Proof of Lemma 2 This lemma follows from the previous Lemma 1. Indeed, from the master equation at $t = 0$, defining $u(0, x) = \hat{u}^{\alpha}(x)$, we have:

$$\partial_t u(t,x)_{|t=0} = \sum_{i=1}^{n}\left(\partial_{x_i}\left[d_{1,i}x_i\hat{u}^{\alpha}(x)\right] + \int_0^{x_i} k^{\theta}_{on,i}(x - he_i)\hat{u}^{\alpha}(x - he_i)c_i e^{-c_i h}dh - k^{\theta}_{on,i}(x)u^{\alpha}(x)\right)$$

$$= \sum_{i=1}^{n}\left(\int_0^{x_i}\left(k^{\theta}_{on,i} - k^{\alpha}_{on,i}\right)(x - he_i)\hat{u}^{\alpha}(x - he_i)c_i e^{-c_i h}dh - \left(k^{\theta}_{on,i} - k^{\alpha}_{on,i}\right)(x)u^{\alpha}(x)\right),$$

where the second equality comes from the fact that $\hat{u}^\alpha$ is the stationary solution of the master equation (18) with burst rates functions $k^\alpha_{on,i}$. Thus, we obtain:

$$\begin{aligned}
\int_\Omega |\partial_t \hat{u}_\alpha(x)|_{|t=0} dx &\leq \sum_{i=1}^n \left(\int_\Omega \int_0^{x_i} \left| k^\theta_{on,i} - k^\alpha_{on,i} \right| \hat{u}^\alpha(x - he_i) c_i e^{-c_i h} dh dx + \int_\Omega \left| k^\theta_{on,i} - k^\alpha_{on,i} \right| u^\alpha(x) \right) \\
&= \sum_{i=1}^n \left(\int_\Omega \left| k^\theta_{on,i} - k^\alpha_{on,i} \right| \hat{u}^\alpha(x) dx \int_0^\infty c_i e^{-c_i h} dh + \int_\Omega \left| k^\theta_{on,i} - k^\alpha_{on,i} \right| u^\alpha(x) dx \right) \\
&= 2\mathbb{E}_{\hat{u}_\alpha} || k^\theta_{on}(X) - k^\alpha_{on}(X) ||_1 .
\end{aligned}$$

Chapter 4

Benchmark with state-of-the-art methods and application to an experimental dataset. Article submitted for publication in Plos Computational Biology.

Those last few years many GRN inference algorithms based on the analysis of single-cell transcriptomic data have been developed. Concurrently, GRN simulation tools have also been proposed for generating synthetic datasets with statistical characteristics similar to the experimental ones. Despite their successes, these two types of methods suffer from severe limitations:

1. GRN inference methods are "top-down" (they do not take into account explicitly the molecular complexity within single-cells) instead of "bottom-up" (based on the molecular processes driving cells dynamics). This makes the inferred interactions difficult to interpret in a mechanistic way, and prevents the possibility of simulating data from the inferred GRN.

2. Simulation methods often need the addition of some adapted technical noise to fit the experimentally-observed variability of gene expression, and the expected patterns generally have to be hard-coded in the model parameters instead of emerging from the GRN mechanistic behavior, preventing their calibration from scRNA-seq data.

In this chapter, we aim to show that using the same mechanistic model (1.16) for both data simulation and network inference allows to overcome these limitations. We already proposed in Chapter 3 a method called CARDAMOM [96] for inferring a most-likely GRN from time-stamped datasets and proved its efficiency on *in silico* datasets generated from small specific networks. In this chapter, we go further in the applications. First, we establish the precision of CARDAMOM as a GRN inference tool, using synthetic datasets generated with HARISSA [38]. Second, we apply the method on a previously published dataset and show that this calibration allows to recover many aspects of the original data when running the model, while providing biological insights into the nature of the detected interactions. For analyzing the results, we use both standard statistical methods and custom criteria based on statistical tests on the marginals of each gene, Wasserstein distances [75] and the reduction dimension algorithms UMAP [62]. These results demonstrate the benefits of using the same model for simulation and inference purposes.

The chapter contains an article which has been submitted for publication in the journal *Cell Systems* [98].

One model fits all: combining inference and simulation of gene regulatory networks

Elias Ventre[1,2,3,6], Ulysse Herbach[4,6], Thibault Espinasse[2,3], Gérard Benoit[1,5], Olivier Gandrillon[1,2,*]

[1]*Laboratoire de Biologie et Modelisation de la Cellule, Ecole Normale Superieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Universite Claude Bernard Lyon 1, 46 allee d'Italie F-69364 Lyon, France.*
[2]*Inria Center Grenoble Rhône-Alpes, Equipe Dracula, Villeurbanne, France*
[3]*Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, Villeurbanne, France*
[4]*Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France*
[5]*Present address; Université de Rennes 1, IGDR - CNRS UMR6290 - Inserm 1305, 35042 Rennes Cedex, France*
[6]*Co-first author*
[*]*Corresponding author: olivier.gandrillon@ens-lyon.fr*

Abstract:
The rise of single-cell data highlights the need for a nondeterministic view of gene expression, while offering new opportunities regarding gene regulatory network inference. We recently introduced two strategies that specifically exploit time-course data, where single-cell profiling is performed after a stimulus: HARISSA, a mechanistic network model with a highly efficient simulation procedure, and CARDAMOM, a scalable inference method seen as model calibration. Here, we combine the two approaches and show that the same model driven by transcriptional bursting can be used simultaneously as an inference tool, to reconstruct biologically relevant networks, and as a simulation tool, to generate realistic transcriptional profiles emerging from gene interactions. We verify that CARDAMOM quantitatively reconstructs causal links when the data is simulated from HARISSA, and demonstrate its performance on experimental data collected on *in vitro* differentiating mouse embryonic stem cells. Overall, this integrated strategy largely overcomes the limitations of disconnected inference and simulation.

Keywords: gene regulatory networks, causal inference, data simulation, transcriptional bursting, stochastic gene expression, single-cell transcriptomics, time-course profiles, lineage commitment

Introduction

Cell decision making as a response to exogenous or endogenous stimuli (e.g., differentiation, proliferation, cell death or biological activity modulation) is often supported by time-dependent modulation of gene expression upon stimulation. Understanding how and why gene expression changes as a function of time in response to specific stimuli is therefore critical to understand the underlying biological processes.

The “how” question can now be approached using single-cell-based technologies, offering an unprecedented resolution and a much finer view than population-based measures [1, 2]. The “why” question relates to the functioning of an underlying gene regulatory network (GRN) which describes interactions between genes through their expression products. GRNs are thus a central notion for understanding and predicting cellular behavior, but their construction from literature is a very laborious task, sometimes even impossible due to the lack of knowledge.

Reconstructing most-likely GRNs from transcriptomic datasets has therefore become a major goal in systems biology [3] but is also notoriously difficult, especially in the case of single-cell transcriptomic data. Indeed, the bursty synthesis of mRNAs, now clearly evidenced [4, 5], gives rise to highly variable and non-Gaussian expression data [1, 6], and current GRN inference methods employ a wide range of statistical and modeling tools [7]. Methods based on a specific mechanistic model have the great advantage of providing biological interpretability, since each inferred interaction between genes can be understood in terms of model behavior. Moreover, such approach generally provides interactions with their direction and intensity, which is not the case for most purely statistical methods.

However, the results of a method based on a mechanistic model can only be considered relevant if the model is able to correctly reproduce single-cell datasets. For instance, it is now widely accepted that the transcriptional bursting phenomenon is associated to specific patterns of gene expression products [8, 9], making continuous single-cell data close to Gamma distributions [10] and discrete data close to negative binomial distributions [11], the latter being themselves mathematically equivalent to Poisson distributions with Gamma-distributed random parameters. Thus, executable network models should at least be able to generate these patterns in their marginal distributions. In any case, the use of a mechanistic model-based method requires prior strong evidence that the underlying model is relevant for simulating realistic single-cell transcriptomic data sets.

We recently developed several methods for inferring GRNs from single-cell data based on a particular mechanistic network model, defined as a ‘multi-agent’ generalization of the well-known two-state stochastic model of gene expression [8] where genes are now being described by interacting two-state models [6]. These methods are well suited for single-cell RNA-seq (scRNA-seq) time-course data, each dataset being considered as a partial observation of the model at a certain time. Crucially, they do not require the observation of cell trajectories, whose inference is a problem in itself [12, 13], but only that the cells sampled at each timepoint are driven by the same dynamical process, i.e., resulting from the same GRN. Our first proposal was called WASABI [14], which uses a divide-and-conquer approach where the problem of GRN inference is solved one gene at a time. Although able to propose relevant GRNs, this approach suffered from two

drawbacks: it required days of computation for a GRN with 50 genes, and proposed a potentially long list of candidate networks. We therefore developed two other methods: HARISSA [6], a GRN simulation algorithm based on the mechanistic model together with a proof-of-concept inference method derived from likelihood maximization, and CARDAMOM [15], a simplified and scalable alternative for the GRN inference part that crucially exploits the notions of landscape and metastability.

In this work, we sought to investigate the benefits of using this model as an integrated tool for both GRN inference and data simulation. We therefore assessed its ability to allow for efficient network reconstruction from time-course scRNA-seq data, while accurately reproducing the dataset main features from the functioning of the inferred network. Note that to the best of our knowledge, this is not performed by existing GRN-based simulation tools, which are generally based on more phenomenological than mechanistic models [16, 17], meaning that gene expression patterns, and especially transitions between cell types, are hard-coded instead of emerging from interactions between genes.

After introducing the setup of our benchmark made from *in silico* datasets generated with the mechanistic model, we first evaluate the performances of HARISSA and CARDAMOM together with four state-of-the-art GRN inference algorithms: GENIE3 [18], PIDC [19], SINCERITIES [20] and SCRIBE [21]. We study the limits of the different categories of inference methods in the case of transcriptional bursting, and verify that the two model-based methods perform better than the others on these datasets. CARDAMOM appears as the best performing algorithm during this benchmark step, which only considers network structures. Importantly, the output of this algorithm is not only a matrix of interaction scores, but also a set of quantitative parameters that can be plugged into the GRN model for simulations.

In a second step, we use CARDAMOM to calibrate the model with a real time-stamped scRNA-seq dataset of differentiating mouse embryonic stem (ES) cells [22]. We demonstrate the ability of the model to reproduce the global features of real time-course transcriptomic profiles. We also show that most of the inferred interactions are indeed supported by biological evidence such as ChIP-seq experiments, although this evidence was not used during the inference process. Altogether, these results establish the ability of an executable network model not only to simulate realistic single-cell datasets, but also to provide an effective reverse-engineering algorithm capable of reproducing the main gene expression patterns of an experimental dataset as emergent properties of the underlying GRN.

Results

HARISSA simulates single-cell datasets from a mechanistic GRN model

We first wanted to benchmark the ability of the different inference algorithms to reconstruct correct network structures from *in silico* generated datasets, i.e., when the ground truth is known. For this, we used the simulation module of HARISSA [23], which generates trajectories of a mechanistic model describing gene expression dynamics (both mRNA and the corresponding proteins) within a single cell, these dynamics being influenced by an underlying GRN and driven

by transcriptional bursting (see Methods and Figure S1). As shown in previous work, this model is indeed able to generate scRNA-seq datasets with realistic marginal distributions [6, 23].

We simulated nine types of datasets corresponding to different network structures (Figure 1): a network of 4 genes with a branching structure and inhibition feedback loop (FN4); a network of 5 genes with a cycling structure (CN5); a network of 8 genes with multiple branching structure and feedback loops (FN8); a network of 8 genes with branching trajectories (BN8); networks with a tree structure of 5, 10, 20, 50, and 100 genes (Trees). These networks represent the main types of network structures that have been used for benchmarking GRN inference algorithms [17]. Overall, the objective was to reproduce time-course experiments in which single-cell profiling is performed after a given stimulus, typically a change of medium [22, 24, 25]. This stimulus was therefore taken into account in all the simulations, in the form of a virtual gene defined as being inactive before the beginning of the experiment and fully activated afterwards.

For each network structure (Figure 1A), the transcriptional bursting model implies that typical single-cell trajectories do not follow a diffusion-like process (at least in the space of mRNA levels), and differ strongly from the more usual and intuitive population-average trajectory (Figure 1B). The practical datasets were obtained by sampling independent cells at a specific sequence of timepoints, therefore not keeping the real cell trajectories but rather considering different cells at each timepoint, forming time-stamped snapshots (Figure 1C). Interestingly, both feedback networks (FN4 and especially FN8) produce a recognizable “differentiation trajectory” across the UMAP space with a clear temporal order of cells. Due to the stochastic nature of cell trajectories generated by the mechanistic model, branching trajectories in snapshots only appear in specific cases, generally when a toggle-switch is dominating the GRN structure and then generates distinct branches in the UMAP representation (see BN8 in Figure 1).

As mentioned previously, HARISSA consists of two modules for performing respectively simulation and inference. Whereas the original inference module of HARISSA was limited to a few genes [6], it recently integrated an effective CARDAMOM-inspired simplification [23] that allows to infer networks with a much larger number of genes. We therefore also benchmarked this method along with the others.

CARDAMOM quantitatively reconstructs causal GRN links

We then inferred GRN structures from the *in silico* generated datasets using the six algorithms presented in the Methods section (HARISSA, CARDAMOM, GENIE3, PIDC, SINCERITIES, and SCRIBE). Note that neither GENIE3 nor PIDC are able to use the temporal information (except for the stimulus state information, which they are also provided with), giving them a disadvantage compared to the other algorithms. They were nevertheless used in the benchmark as they are considered to be among the best algorithms for single-cell data, and given that very few algorithms are specifically adapted to time-stamped datasets. Indeed, most methods are limited to static data, and those that are not (such as SCRIBE) require temporally-ordered cell trajectories instead of independent snapshots, thus requiring a pre-processing step that can itself be subject to errors. Moreover, it was not known how they would fare in a time-course setting with transcriptional bursting, which was an interesting question per se.

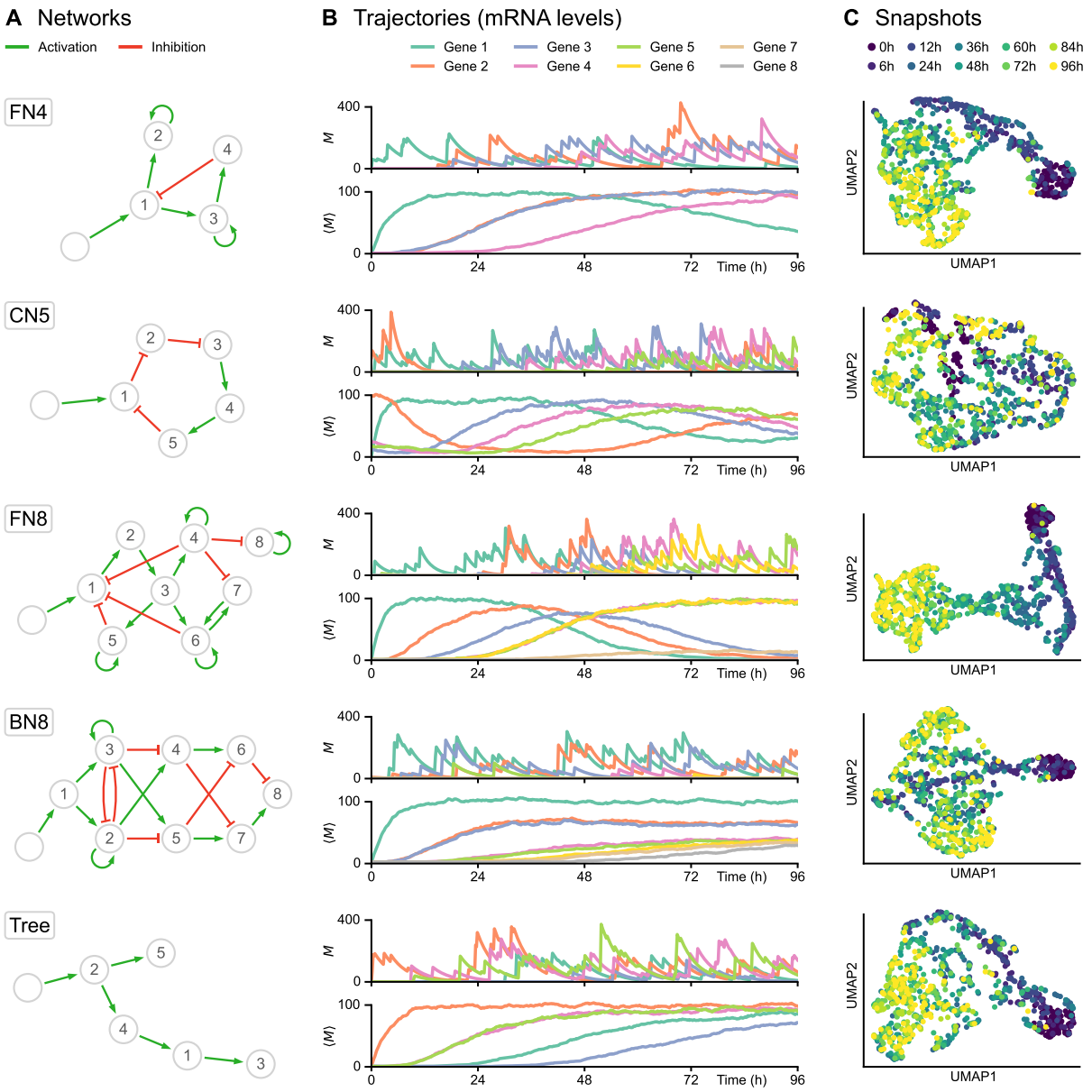


Figure 1: Single-cell data simulation using HARISSA. (A) Networks used for subsequent tests, including feedback loop networks (FN), a cycling network (CN) and a branching network (BN). Genes and stimulus are represented by numbered nodes and an empty node, respectively, while green arrows indicate activation and red blunt arrows indicate inhibition. **(B)** Corresponding trajectories, defined as time-dependent mRNA levels (in copies per cell). For each network, the first plot shows one example of single-cell trajectory M while the second plot shows the population average $\langle M \rangle$ from 1000 cells. The transcriptional bursting model underlying HARISSA implies that every single-cell trajectory differs strongly from the more usual population average. **(C)** Two-dimensional UMAP representations of corresponding single-cell snapshots, defined as mRNA levels sampled at 10 timepoints in different cells from 0h to 96h, with 100 cells per timepoint. Such snapshots are called *time-stamped data* in the text and are fundamentally different from single-cell trajectories, which are currently not available experimentally.

We also emphasize that among these algorithms, only CARDAMOM and HARISSA have the significant advantage of providing biological interpretability, thanks to the mechanistic model on which they are based: here the network parameters are not mere interaction scores, but quantitative parameters that can be plugged into the model for simulations. From this perspective, even similar performances compared to the other algorithms would be satisfying.

Inference was performed on ten independent datasets for each condition, and the results were merged into the area under the precision-recall curve (AUPR) which measures the quality of the inferred GRN structure. We also compared the inferred GRNs with a naive method consisting in assigning to each edge of the network the value given by the Pearson correlation coefficient between the corresponding genes (abbreviated as PEARSON): this comparison with Pearson coefficients makes it possible to verify, when the algorithms show good performances, that these are not only due to highly correlated data which are thus not difficult to analyze. The results are presented in Figure 2A-B for the first five algorithms. We present the results for SCRIBE separately in Figure 2C because this algorithm requires temporally or pseudo-temporally ordered trajectories, and the results then depend on the pre-processing that is applied on the time-stamped data.

CARDAMOM and HARISSA appeared to outperform the other algorithms for most of these datasets. In particular, in terms of directed interactions, these two methods always clearly performed better than the others. The undirected networks for which GENIE3 and PIDC have similar performances (CN5 and BN8) correspond to cases where the Pearson correlation method is also accurate.

Also, if GENIE3 and PIDC represent an improvement over the Pearson correlation method, they seem to perform poorly when the correlation between genes is not sufficient to infer a reliable GRN. More precisely, we observe that GENIE3 and PIDC are accurate for tree-like networks (Trees), even with bifurcating trajectories (BN8) and cycling (CN5), which was not the case in [17]. On the contrary, SINCERITIES performs very poorly for these type of networks, but seems however competitive for networks with feedback loops (FN4 and FN8) where GENIE3 and PIDC have lower performances. These networks are more difficult to reconstruct. Indeed, as visible in Figure 1, the population-average trajectories of some genes are completely similar. Some genes also have the same marginal distribution of mRNA levels: for example in the network FN4, gene 2 and gene 3 have the same input (gene 1), so their marginal distributions evolve similarly at each timepoint. Then SINCERITIES, which bases the inference procedure on the approximate distribution for each gene, fails to make this subtle distinction, illustrating the improvement that is typically expected from HARISSA and CARDAMOM. On all the networks, GENIE3 fails to infer reliably the direction of the interaction, i.e., to distinguish the interaction $i \to j$ from the interaction $j \to i$. On the contrary, because of their mechanistic assumptions, CARDAMOM, HARISSA and SINCERITIES have always quite similar results for directed and undirected inference. Finally, we observed that CARDAMOM outperforms HARISSA on most of the networks.

Regarding SCRIBE, we tested its performances for 3 types of data (Figure 2C):

1. When we have access to real trajectories (*Real traj.*): each cell at each timepoint is being associated to a real ancestor at the previous timepoint and a real descendant at the

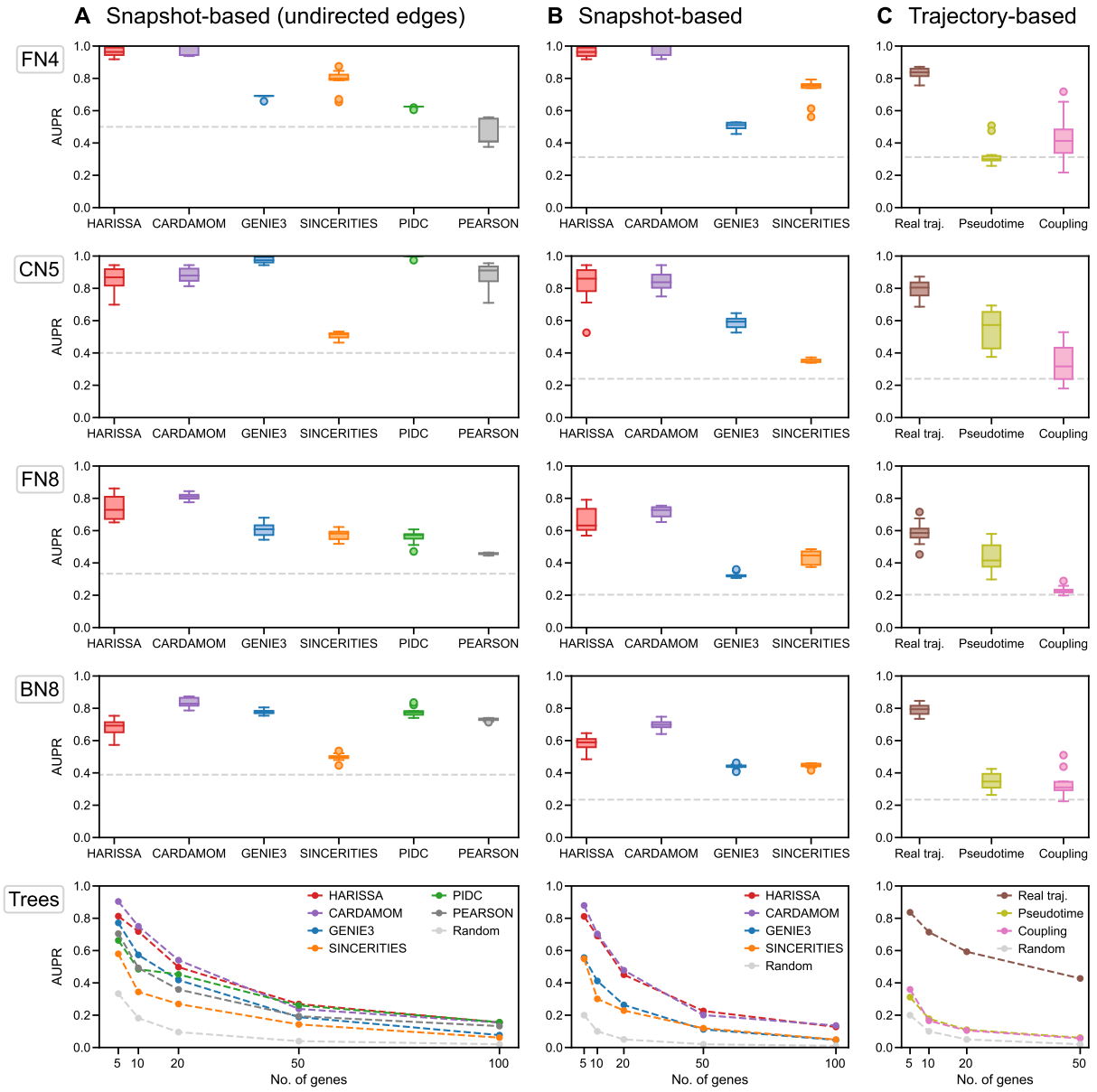


Figure 2: Benchmark of inference methods for five different network structures. For each network, inference is performed on ten independently simulated datasets, each dataset containing the same 10 timepoints with 100 cells per timepoint (1000 cells sampled per dataset). The performance on each dataset is then measured as the area under the precision-recall curve (AUPR), based on the unsigned inferred weights of edges. Finally, the performance of each method is summarized as a box plot of the corresponding AUPR values, or the average AUPR value for the tree-structure activation networks (*Trees*). For each plot, the dashed gray line indicates the average performance of the random estimator (assigning to each edge a weight 0 or 1 with 0.5 probability). For the *Trees* networks, each dataset corresponds to a random tree structure of fixed size (5, 10, 20, 50, and 100 genes) sampled from the uniform distribution over trees of this size. **(A)** Performance of all methods when considering only undirected interactions. **(B)** Performance of the methods able to infer directed interactions. **(C)** Performance of the SCRIBE inference method for the same networks, in three conditions: when one has access to real single-cell trajectories (in brown), when pseudo-trajectories are reconstructed from time-stamped data using a coupling method similar to Waddington-OT (in pink), and when a single pseudo-trajectory is reconstructed using the pseudotime algorithm SLINGSHOT (in light green).

following one. Such knowledge can of course only be accessed with *in silico* generated datasets or *in vitro* for a very limited number of genes by using live-cell imaging of short-lived transcriptional reporters [26];

2. When we do not have access to real trajectories, and each cell at each timepoint is associated to a pseudo-ancestor at the previous timepoint and a pseudo-descendant at the following one, using the Waddington-OT method described in [27] (*Coupling*);

3. When we do not have access to real trajectories, and the algorithm SLINGSHOT [28] is used for reconstructing a pseudo-temporally ordered trajectory (*Pseudotime*).

We observed that SCRIBE performs well when the datasets are built with the method 1, but poorly when the datasets are built with the methods 2 and 3, at least on the tested networks (Figure 2C). These poor performances are due to the loss of temporal coupling between measurements of genes that interact. They suggest that neither optimal coupling nor pseudotime reconstruction are sufficiently efficient for GRN inference in case of transcriptional bursting. Concerning the optimal coupling method, we notice that this might be due to the "movement by diffusion" assumption on which the method Waddington-OT is built, which does not takes into account the constraints on the trajectories imposed by the GRN.

When computing the average runtime of each algorithm on the tree-like networks, we observed that except for SCRIBE, all algorithms are suitable for inferring GRN with a realistic number of genes (see Table S1). Thus, due to this computational limit and its poor performances when using time-stamped data, we did not consider SCRIBE for further analysis.

We then investigated the limit performances with respect to the number of cells and/or timepoints. We observed that the performances of the first five algorithms decrease for the tree-like networks when the number of genes increases (Figure 2). This can be due to three main factors:

1. A sequence of timepoints too coarse in relation to the dynamics would directly lead to a lack of inference accuracy;

2. A sequence of timepoints which is too restricted may not allow to see interactions involving some genes that are regulated late in the process. For example, in Figure 2, we observe that the inference on the Trees networks is very poor for more than ten genes: it comes from the fact that some genes are never activated before 96h;

3. The number of cells at each timepoint can simply not be enough to infer a reliable GRN.

We therefore investigated the effects of these three factors on the accuracy of the algorithms by studying their performances in terms of AUPR for ten datasets generated from ten randomly-generated tree-like network of ten genes, when varying the number of cells at each timepoint (Figure 3A), the length of the interval for a fixed time gap between each timepoint (Figure 3B), and the density of the sequence of timepoints for an interval with fixed length (Figure 3C). As anticipated, all these conditions have an impact on the quality of the inference: augmenting their values tends to produce a better quality of inference. We also observed that the number of sampled cells seems less critical than the other factors, confirming that few cells at a sequence of timepoints which is dense and long-enough is preferable to many cells on a sequence of

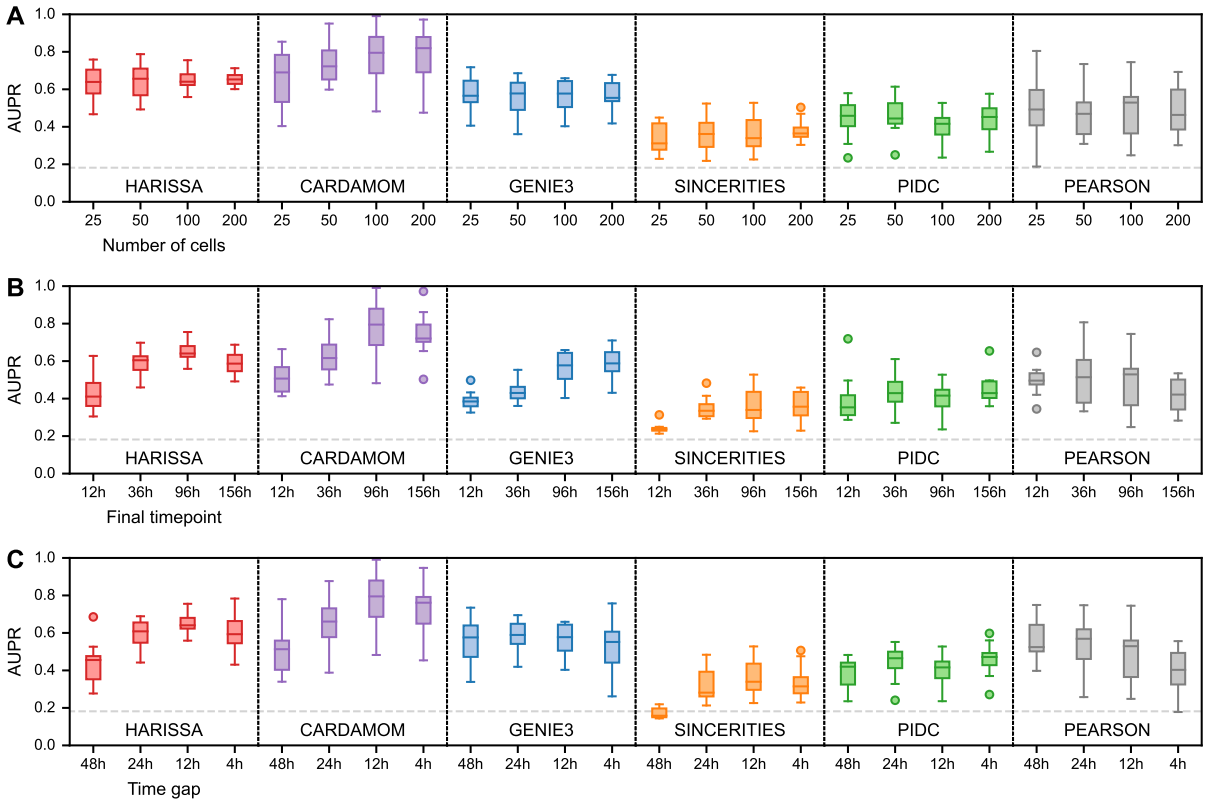


Figure 3: Dependence of inference methods on data collection parameters. For simplicity, only the case of undirected interactions is considered here and the datasets are restricted to 10-gene tree-structure networks (see Figure 2 for the general benchmark). Inference is performed for each method and condition on ten independently simulated datasets and summarized by box plots of AUPR values as in Figure 2. **(A)** Performance as a function of the number of cells per timepoint, while keeping the same timepoints. **(B)** Performance as a function of the length of the measurement period, while keeping the same gap between timepoints and the same total number of cells. **(C)** Performance as a function of the density of the measurements, while keeping the same final timepoint and the same total number of cells.

timepoints which is too coarse and/or too short. This should be kept in mind when designing single-cell transcriptomics experiments aiming at GRN inference.

Hence, both CARDAMOM and HARISSA, with a benefit for using CARDAMOM, allowed to efficiently reconstruct network structures by reverse engineering the generative model on which they are based. We then needed to test its ability to reproduce an experimental dataset from the literature after network inference.

Application to a real dataset yields a biologically relevant network

As a test case, we used a time-stamped *in vitro* dataset from Semrau et al. [22] obtained by scRNA-seq of a retinoic acid (RA)-induced differentiation of mouse ES cells (see Methods). This well-characterized model system of *in vitro* differentiation recapitulates the transition from pluripotent embryonic stem cells towards two cellular lineages (ectoderm- and extraembryonic endoderm-like cells), all characterized by well-established molecular markers that were further used in GRN inference.

In order to interpret the resulting GRN, we sought to assess whether the inferred interactions

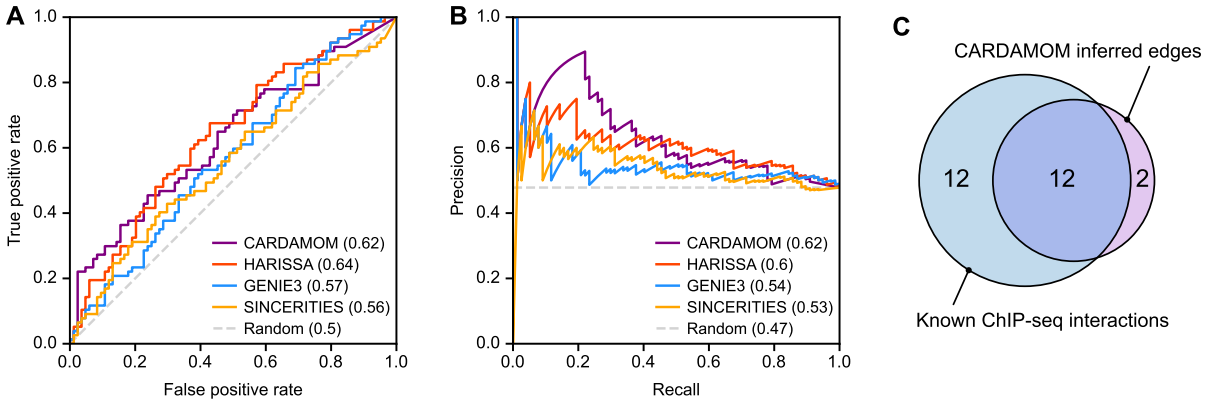


Figure 4: Comparison between inference methods and physical interactions derived from ChIP-seq. The four directed GRN inference methods were applied to the experimental dataset from [22] restricted to a panel of 41 marker genes identified by the authors, and a reference network was obtained independently from edges supported by ChIP-seq data. As we only have access to physical interactions involving the retinoic acid (RA) stimulus or genes Pou5f1, Sox2, and Jarid2, the comparison only considers the related edges. **(A)** Receiver operating characteristic (ROC) curve and corresponding area under the curve (AUROC) for each inference method. **(B)** Precision-recall (PR) curve and corresponding area under the curve (AUPR) for each inference method. **(C)** Venn diagram showing the overlap, for interactions involving the RA stimulus, between directed edges predicted by CARDAMOM and known physical interactions identified by ChIP-seq analysis.

are supported by known biochemical evidence of physical interaction between regulators and regulated genes (Figure 4). For this, we annotated the inferred edges coming from genes encoding known transcription regulators (i.e., transcription factors and cofactors) included in the network and for which ChIP-seq data are currently available in ES or the closely related embryonic carcinoma (EC) cell system. Since the RA stimulus exerts its differentiating effect mainly through the members of the RA-activated nuclear receptors subfamily RAR (NR1B) that encompass 3 paralogs (i.e., RARα, RARβ and RARγ), the annotation of the interaction edges linking the stimulus and the regulated genes was based on the presence/absence of ChIP-seq peaks for any RAR paralogs at less than 10 kb upstream or downstream of the annotated transcription start site (TSS) in RA-stimulated ES or EC cells [29, 30]. Although arbitrary, the chosen distance between TSS and DNA binding site for the indicated transcription regulator is relatively conservative as transcriptional effect could be exerted from greater distance up to megabases [31] and the absence of supporting peak as defined should not be interpreted as a proof of absence of any direct modulating effect. Similarly, the edges supported by physical interactions data for Sox2 and Pou5f1, or Jarid2 were extracted from [32, 33].

Using these known physical interactions as a ground-truth, we compared the receiver operating characteristic (ROC) and precision-recall (PR) curves related to the network structures inferred by the four algorithms (Figure 4A-B). We observed that, in accordance with the previous results, CARDAMOM and HARISSA appear as the top-ranked algorithms, displaying both a very close ability to infer known edges.

We then examined the structure of the network inferred by CARDAMOM (Figure 5). Importantly, in agreement with its differentiating effect in ES/EC cell systems, we observed that the RA stimulus is densely connected with genes involved in pluripotency maintenance as supported by multiple

biological analyses [29, 30, 34] and to a lesser extent with gene nodes corresponding to genes associated with specific cell fates, this latter observation likely reflecting how the stimulus is modeled (see Methods). Notably, these last nodes also exhibit a relatively high interconnectivity (e.g., endodermal differentiation) as compared to intergroup connectivity. Although biologically interesting, these observations illustrate our previous conclusion and likely mirrors the unbalanced experimental design characterized by a dense sequence of timepoints during the early phase (0h to 36h) of the differentiation process analysed and a coarser sequence of timepoints in the mid (36h to 48h) and late (48h to 96h) phases of the process [22].

Most notably, the overwhelming majority (0.85%) of the inferred edges that involve the RA stimulus are supported by biochemical evidence (Figure 4C). Similarly, the edges inferred from Pou5f1, Sox2, and Jarid2 nodes are globally supported by physical interaction (2/3 for Pou5f1, 2/3 for Sox2, and 1/1 for Jarid2).

We also observed that some inferred edges are not supported by documented physical interactions, as expected for genes encoding proteins unable to directly interact with DNA. As an example of such node, Sparc (also known as Osteonectin) appears highly connected to genes associated with all four cell states despite its inability to directly interact with gene basal transcription machinery (i.e., RNA polymerase complex). However, the inferred edges are clearly in agreement with its documented role in promoting endodermal differentiation [35]. Additionally, unsupported inferred edges may mirror the lag time between the expression and therefore the physical interaction between regulator and regulated genes and the observed transcriptional effect. By contrast with TFs that establish contact with the transcription machinery, modifying cofactors often catalyse deposition/erasure of epigenetic marks (e.g., acetylation/methylation of histones, DNA methylation) that will likely modulate transcription in a longer lasting manner. In this respect it is interesting to note that the Dnmt3b gene negatively interacts with many other genes in the network, which mirrors the fact that Dnmt3b is a *de novo* DNA methyltransferase, and has an indirect effect on gene regulation through CpG methylation, a well documented epigenetic mark generally associated with gene expression silencing. Altogether, this illustrates that our GRN model does incorporate various epigenetic information or indirect effects and is not restricted to physical interactions between transcription factors and their target genes.

While most inferred edges involving genes that encode TFs appear to be supported by physical interactions, many physical interactions detected by ChIP-seq are missed by CARDAMOM (e.g., 50% for RA, see Figure 4C). This observation is however not necessarily the sign of a lack of accuracy of the inferred GRN, since the detection of a physical interaction is not *per se* the hallmark of a modulating effect on the transcription level of the target gene [30]. Additionally, some specific regulatory structures are notoriously difficult to infer as illustrated by the high failure rate (96%) in inferring edges from Jarid2, a component of a repressive complex expressed in the pluripotent state and directly involved in the silencing of differentiation-associated genes. Interestingly, the interaction between Jarid2 and most of its physical targets presented in our GRN were instead wrongly detected as an inhibitory effect of the regulated genes on their regulators. This is due to the fact that CARDAMOM works by going forward in time, and thus fails to capture an inhibition that has an effect at the beginning of the process and which can be detected only further: instead, it would be prone to interpret the increase of the repressed genes by the effect

of other intermediate genes, and the decrease of the repressor by an inhibitory effect of the repressed genes. We discuss further such bias of the algorithm in the Discussion section.

For a better understanding of the inferred GRN dynamics, we also examined a dynamical network representation, where each edge appears at the timepoint for which it was detected with the strongest intensity by the inference algorithm (Figure 6). Unsurprisingly, the RA stimulus is detected at the earliest timepoint of the response (6h) and then ceases to influence the signal, which propagates in waves through the network as we described in a previous study [14]. For example, we do clearly observe the late increase of interactions for genes involved in specifying the extraembryonic endoderm.

Simulation of the inferred network reproduces the original dataset

While inferring the GRN structure, CARDAMOM also inferred all the other parameters of the model, as described in Methods, except the mRNA degradation rates $d_{0,i}$ for each gene i. These parameters $d_{0,i}$ are not negligible as they scale the dynamics of the process. To address this problem, we used values from the literature that can be found in [36] (see Methods and Table S2). Once the model has been calibrated, one can simulate an *in silico* dataset and sample the nine timepoints corresponding to the *in vitro* experiment. More precisely, we simulated two different datasets after calibrating the mechanistic model: one actually using the inferred network interactions, and one corresponding to the "null network" defined by removing the interactions (i.e., all genes individually calibrated with the same parameters but kept independent).

We first decided to verify, as advocated by Soneson et al. [37], the suitability of our generated synthetic data. We used countsimQC [37], a recent tool for comparing multivariate single-cell datasets, already used for benchmarking synthetic scRNA-seq data [38] (see Methods for more details). The synthetic dataset indeed mimics experimental data for a large number of tested characteristics (Figure S2). However, we observed that except for correlations, the features considered by countsimQC are also well reproduced by the dataset simulated with the null network. This suggests that countsimQC features are not sufficient for measuring the accuracy of the dataset reproduction.

We then explored the ability of the synthetic dataset to recapitulate more sophisticated dimensions of the experimental data. At that stage, the critical question concerns the temporality of the synthetic data. Indeed, as developed in [15] and [23], none of the algorithms used for the benchmark (and presented in Methods) allows to take into account the real temporality of the data: the only information they use is the order of the sequence of timepoints at which the cells are measured. It is therefore not necessarily expected that a dataset simulated with the network inferred by CARDAMOM can reproduce the data distribution exactly at the same timepoints. The temporality is taken into account in a second time, by setting the value of the degradation rates from the literature. However, the hypothesis that these degradation rates are not time-dependent (which of course oversimplifies the biological reality [39]) may prevent us from being able to perfectly fit the time-dependent evolution of the data.

We observed that this hypothesis indeed limits our ability to simulate the true dynamics at the last timepoint. In particular, the process seems to accelerate between 72h and 96h and the model

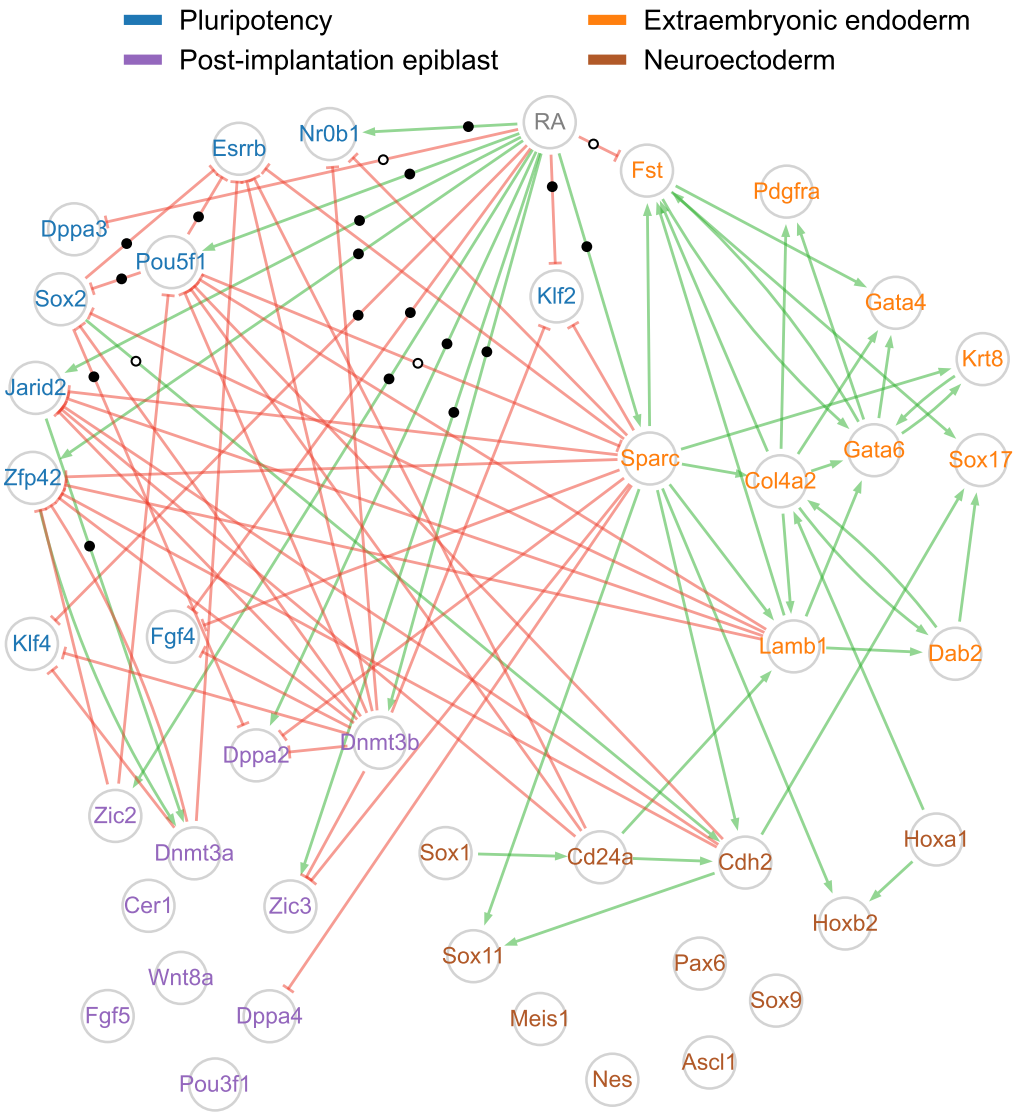


Figure 5: Network inferred by CARDAMOM from a real time-stamped scRNA-seq dataset. The CARDAMOM inference method was applied to the experimental dataset from [22] restricted to a panel of 41 marker genes identified by the authors. The network structure is obtained by keeping only the 5% strongest activations (green arrows) and inhibitions (red blunt arrows) acting on each gene. Genes are colored according to four groups related to different cell states (pluripotency, post-implantation epiblast, neuroectoderm, extraembryonic endoderm) following the proposed classification of [22]. Edges supported by a ChIP-seq interaction are marked with black dots (see main text for the definition of what is considered as an interaction) and edges that are not supported are marked with white dots. Edges for which we have no reliable information have no mark.

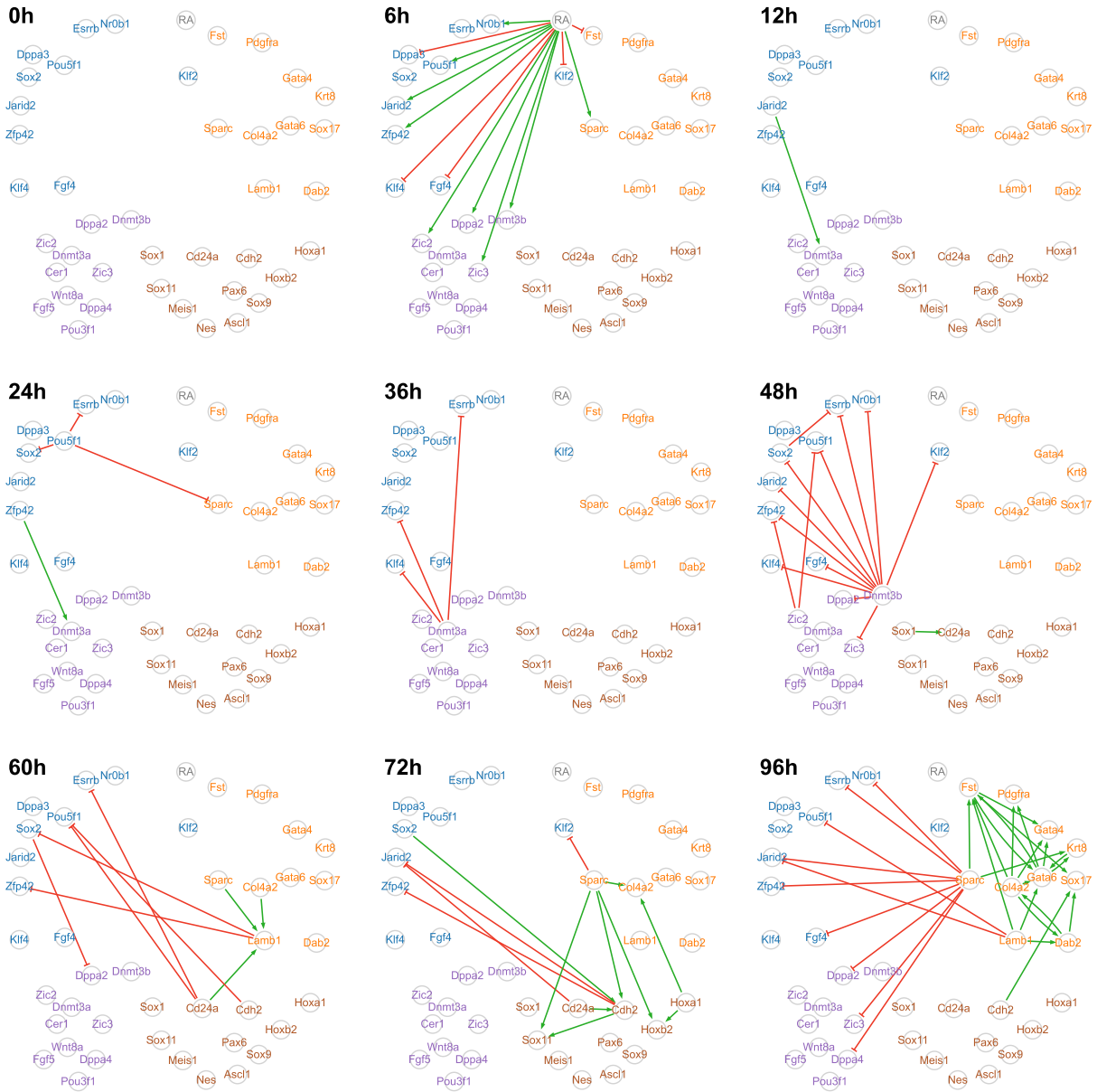


Figure 6: Time decomposition of the network inferred by CARDAMOM from a real time-stamped scRNA-seq dataset. Decomposition of the network shown in Figure 5, where each edge appears at the timepoint for which it was detected with the strongest intensity. This dynamic representation highlights a consistent flow of information coming from the stimulus. Gene positions and colors as well as activation and inhibition representations are the same as in Figure 5.

cannot be in adequacy with both the dynamics between 0h and 72h, and between 72h and 96h with the same degradation rates. It is important to note that such a global variation in degradation rates has been observed experimentally during the differentiation of chicken erythroid progenitors (see [24] and data at https://osf.io/k2q5b/). We thus decided to multiply the degradation rates by a scaling factor after 72h, allowing the process to reach its final state in time.

We compared these datasets using different metrics. First, we examined the extent to which the simulations matched the experimental marginal distributions of each gene. In Figure 7A, we represent the time-dependent evolution of the p-value of a Kolmogorov-Smirnov test between the GRN-generated distributions and the experimental dataset. We can see that some genes are better fitted than others, as exemplified by the Esrrb gene that is correctly fitted while the Sparc gene seems more difficult to catch (see Figure 7B). We nevertheless observe that for most genes and timepoints, the p-value is above 5%, meaning that the marginal distributions of the experimental data are quite well reproduced by the GRN model. We compare in Figure 7C the mean Earth Mover Distance (EMD), and in Figure 7D the mean p-value of the Kolmogorov-Smirnov test applied on the 41 genes at each timepoint, between the empirical distributions of the experimental dataset and the two simulated datasets (with and without GRN). We observe that without GRN, the distance between the distributions generated by the model are constantly increasing, that is diverging from the experimental datasets (Figure 7C). This is corroborated by the fact that the mean p-values are decreasing monotonically (i.e. the model's output is more and more significantly different from the experimentally observed distributions, see Figure 7D). The behavior of the GRN-simulated dataset is much closer to the experimental one, as seen from a smaller (and constant) EMD distance as well as a larger mean p-value. However, we observe in Figure 7A that the mid-timepoints (the central portion of the dynamics) seems to be the most difficult to capture, since mid-time p-value are often higher that at the beginning or end of the kinetics. This is corroborated by Figure S3, where we plotted the temporal marginal distributions of six genes, and where Sox2 and Sparc in particular appear to have a final distribution close to the experimental one but not the correct transient behavior.

Finally, we were interested in how well we could capture the joint distributions. For this, we compared UMAP representations of the experimental dataset (Figure 7E) and the datasets simulated from the inferred network (Figure 7F) and from the null network (Figure 7G). These three datasets were projected on the same pseudo-axes based on the UMAP computed from the experimental dataset, using the methodology described by McInnes et al. [40]. This common projection allows a better side-by-side comparison between datasets. It is immediately evident that our GRN-generated data points are very closely mimicking the actual experimental data points, and that this resemblance is completely lost if all interactions are removed. The fact that UMAP is not linear requires some precautions, that we discussed in the Methods section.

To conclude, we observed that the mechanistic model can reproduce the major characteristics of the gene expression patterns observed during a differentiation process examined at the single cell level. It also appears clearly that simulating the model with the network inferred by CARDAMOM significantly improves the fit to the experimental dataset compared to the simulation with the null network (see Figure 7 and Figure S5).

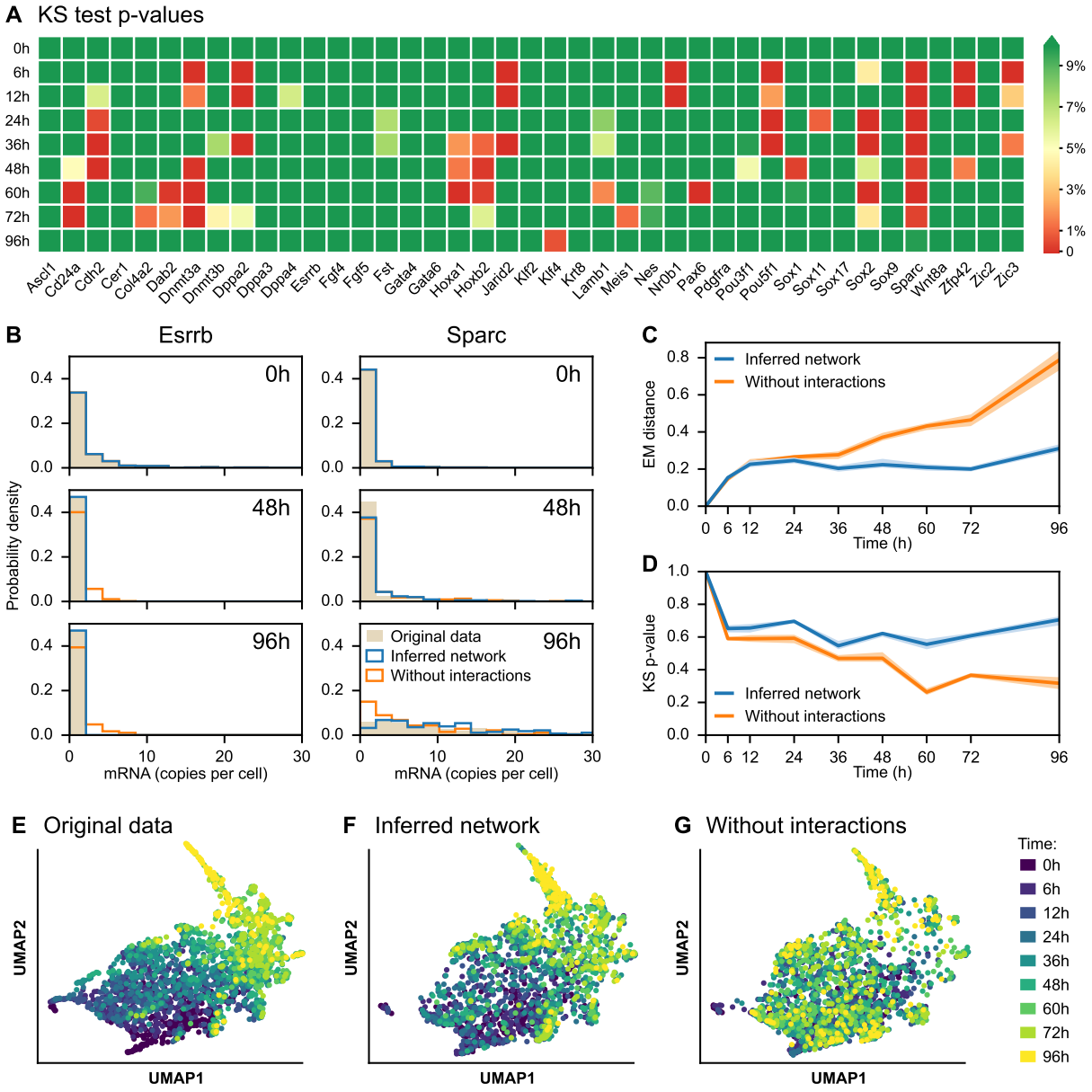


Figure 7: Inferred network simulations compared to the original dataset. (A) Heatmap of p-values associated with Kolmogorov–Smirnov (KS) tests between real and simulated mRNA distributions, for each of the 41 genes of the network and for each timepoint. The green color indicates p-values greater than 5%, implying that the model output is not significantly different from the experimental dataset. **(B)** Time-dependent distributions of Esrrb and Sparc genes for the experimental dataset (*original data*, in beige) and datasets simulated after calibrating the mechanistic model, one including interactions (*inferred network*, in blue) and one obtained after removing interactions (*without interactions*, in orange). **(C-D)** Average earth's mover (EM) distance **(C)** and average KS p-value **(D)** between real and simulated distributions, for the inferred network and without interactions. The dispersion corresponds to the first and ninth decile from ten simulations. **(E-F-G)** Two-dimensional UMAP representations of the original dataset **(E)** and the datasets simulated from the inferred network **(F)** and without interactions **(G)**. In these three plots, Sparc is removed from the genes represented in Figure 5 as its dynamics are not well captured by the mechanistic model, so the three datasets consist of 2449 cells with mRNA levels of 40 genes.

Discussion

The major interest of the method we proposed in this work is that it uses the same model for both inferring and simulating a gene regulatory network. The simulation part is not a novelty: a growing number of algorithms are proposed for simulating realistic single-cell gene expression datasets [38], and some have already been used for benchmarking GRN reconstruction methods [16, 17, 41].

Part of the success of our GRN model lies in its ability to reproduce the main characteristics of cell-cell heterogeneity observed in experimental datasets by making stochasticity an inherent part of the model, instead of adding an adapted noise a posteriori (see Methods). This is not the case of most algorithms that are used to simulate gene expression data associated to a regulatory network for benchmarking purposes [42, 43], even when they are based on underlying mechanistic models [16]. This point is also crucial for developing analytical results able to be used for the reverse engineering of the model, as it has been done for HARISSA and CARDAMOM.

Among the recently described algorithms, SERGIO [16] is the closest to our work. Nevertheless, we want to highlight some key differences:

1. SERGIO's mechanistic model is based on SDEs, treating noise as a Gaussian white noise. This is clearly insufficient to capture the *biological zeros* and Gamma-shaped variability, which in SERGIO arise only from the addition of technical noise. In our modeling scheme, these features arise naturally from the transcriptional bursting phenomenon.

2. As stressed by the authors, SERGIO seeks to simulate data with an explicit GRN as an input, a forward simulation goal, rather than attempt to estimate it from data, a reverse engineering goal. This is a fundamental difference with our work where we seek *simultaneously* to infer and to simulate data. To the best of our knowledge, the ability to do so using a mechanistic model is a true novelty of our work.

3. Our modeling scheme is amenable to in-depth mathematical analysis [6, 15, 23, 44].

4. The use of a specific module to add technical noise, as well as SERGIO's ability to generate both spliced and unspliced versions of mRNAs are welcome innovation that we will consider in future versions of our work.

We also mention that there has been a recent surge of interest in using generative adversarial networks (GANs) for producing realistic new single-cell transcriptomics data [45]. Although it can be an efficient strategy for data generation and augmentation, its behaves as a black box regarding the underlying biology. Our main added value here is that our model is based on the biophysical reality of the cell and provides a clear materialistic explanation for generated data. Since we have shown that this generative model can be calibrated from single-cell datasets, it can also be used for control purposes, aiming at controlling the cellular phenotype by interfering with the GRN behavior.

While the test case was made using a dataset obtained during a differentiation sequence, one should note that our approach can be applied to any biological process for which time-stamped single-cell transcriptomic data are obtained after applying a given stimulus. When

such time-stamped snapshots are not available, the algorithm could in principle take as an input time-reconstructed data (i.e., artificially ordered snapshots). In that case, the quality of the inference will strictly depend on the effectiveness of the time reconstruction algorithm.

Although efficient and promising, the model and the method we presented here have some limitations that are clearly identified, and that should guide future research efforts. First, we consider that the burst frequencies, which are critical parameters of the regulation, are sigmoid functions of protein levels. This implies that the distribution of a gene generally cannot have more than 2 modes [15], one associated to a low frequency of bursts and one to a high frequency of bursts, which correspond to the regions of the gene expression space where the sigmoid is relatively flat. Thus, if the distributions of some genes appear more complex than a mixture of these 2 modes, the model is not expected to reproduce accurately its dynamics, since some minor regulatory interactions should be not detected (in particular if the "hidden" mode is much closer to one of the two main modes than the other). This seems to be the case for example for the slight decrease of Sparc at t = 24h, that would have been better capture by adding an extra mode. To solve these kind of errors, it could be necessary to complexify the model, for example by modeling the burst rate functions by a multi-layer perceptron rather than a sigmoid as proposed in [15].

Second, CARDAMOM uses the temporality linearly, by taking into account the timepoints one after the other without possibility of backward step. This explains why the algorithm is unable to detect the known fact that a gene like Jarid2 inhibits some Extraembryonic and Neuroectoderm genes: indeed these inhibitions could only be detected by going up the arrow of time when the target genes see their expression increase, in order to find the real cause of this effect. Instead, we have seen that the algorithm interprets it as an activation of some other genes. We believe that it could be tackled by taking inspiration from the Recurrent Neural Network (RNN) theory, but it should be achieved while keeping the interpretability of the results. Note that we already developed a similar analogy for the regression step at each timepoint, which can be interpreted as the learning step of a perceptron [15].

From that point of view, if the information was indeed transmitted forward, the network would be supposed to be completed step by step until reaching its complete form. However, two types of incompatibilities may still occur:

1. A direct incompatibility occurs when an edge which has been inferred at a certain timepoint is chosen to be reset to a value close to 0 or even to change sign at another one.

2. An indirect incompatibility corresponds to the case where the effect of an edge $(j \to i)$, which has been inferred at a certain timepoint, is compensated by another edge $(k \to i)$ at a following timepoint, but that the gene k expression products were high enough at the previous timepoint to thwart the effect of the relation $(j \to i)$.

This explains why, for the experimental dataset presented in the Results section, the model is not able to reproduce the behavior of some genes. One good example is Pou5f1 between t = 0h and t = 24h: indeed, the edges that could generate the slight increase of Pou5f1 at t = 6h are thwarted by the edges inferred at the following timepoints which lead to the strong decrease of Pou5f1 after t = 48h. However, these incompatibilities could also have a biological meaning, and

be impossible to solve by modifying the regression problems. To go further, it is necessary to discuss the notions of structure and states of the network, which is related to the importance of possibly hidden variables. If the network incorporates all critical nodes, then the structure should not change. But it is different if there are some hidden variables, like genes the level of which are not measured, which results in modifying the network structure. For example, the problem of Pou5f1 that we have presented above could be explained in the following way: at t = 0h, an hidden variable may act on the interaction by preventing its possibility, before the hidden variable disappear at t = 24h. So if the hidden variable was integrated, then the network structure would not change anymore, but it is likely that we should in our case consider a modification of the network structure at t = 24h.

Third, the model does not allow the synthesis and degradation rates to vary over time. We have already mentioned that this was a problem for simulating the passage from t = 72h to t = 96h, and we decided to speed up the last time step for the model to reach its stationary distribution at 96h. We believe that most of the errors observed in the simulation with respect to the experimental dataset could be solved by finding an appropriate degradation rate. Thus, a significant improvement would consist in taking into account at each regression step the size of the time interval, and not only its order in the series of timepoints as it is now the case. This could allow to find a most-likely GRN in accordance to the degradation rate at each timepoint, or even to infer a most-likely degradation rate for each timepoint. Note however that if the latter case could both provide us a new information on the variation of degradation rates during differentiation and a better accuracy on the relative importance of the interactions in the GRN, it could also accentuate the problems of identifiability, and should therefore be studied carefully.

Fourth, the model does not take into account proliferation nor apoptosis while studying the stochasticity of the differentiation processes, nor the regulation of the proliferation rate by gene expression products. When sampling a distribution of n cells at a time t, the initial condition is built by sampling n cells under the uniform distribution among the set of cells at 0h, and to simulate its evolution during t hours. However, if some cells are supposed to have a higher death rate, and others to have a higher division rate, the process should evolve preferably in a certain direction in the gene expression space, which is going to be ignored in the current version of CARDAMOM. Taking into account these characteristics is a notoriously difficult task: a significant improvement has been recently achieved with Waddington-OT [27, 46], where a stochastic diffusion process models gene expression dynamics. Extending this kind of approach for the mechanistic model will be the subject of future works.

Finally, future versions of our method may consider additional biological features such as spatial cell-cell communication as the advent of multiomics datasets should provide data allowing to analyze the effect of these processes on differentiation, which is not possible in the case of scRNA-seq data without seriously compromising identifiability. We believe that the work presented here could serve as a basis for developing multiscale approaches to differentiation processes.

Methods

Mechanistic model of gene regulatory networks

The model used throughout this article is based on a hybrid version of the well-established two-state model of gene expression [6], where a gene is described by the state of a promoter, which can be either *on* or *off*. If the promoter is *on*, mRNAs are being transcribed at a rate s_0, which are then translated into proteins at a rate s_1. Degradation of both mRNAs and proteins occurs at a rate d_0 and d_1, respectively. The transitions between the on and off states occur at times of rates k_{on} and k_{off}. We consider the *bursty* regime of this model ($k_{\text{on}} \ll k_{\text{off}}$), corresponding to short active periods with high transcription rates, as experimentally observed [47–50]. In this regime, mRNA is then transcribed by bursts of tens to hundreds of molecules. The random times at which these bursts occur are still described by an exponential distribution of parameter k_{on}, and their random size by an exponential distribution with mean s_0/k_{off}. This model is compatible with experimental single-cell data, as steady-state mRNA levels follow for each gene a Gamma distribution, in line with continuous single-cell data [10].

The key idea is to incorporate this model into a network: the burst rate for each gene i is given by a gene-specific function $k^{\theta}_{\text{on},i}(P)$, where P is the vector of protein quantities (Figure S1). This function depends on proteins through a GRN, represented by an n-by-n matrix $\theta = (\theta_{ij})$ where n is the number of genes in the network. The value of $k^{\theta}_{\text{on},i}(P)$ then corresponds to the transcriptional burst frequency of gene i given protein levels P. Each parameter θ_{ij} encodes the interaction $j \to i$ with its direction, sign, and intensity. Recent work suggests that burst sizes are smaller and more uniform than previously anticipated [49] therefore leaving more room for burst frequency modulation [51] as a mechanism for gene expression regulation. We therefore consider that interactions come mainly from the modulation of burst frequencies $k^{\theta}_{\text{on},i}$ and that for any gene i, the rates $k_{\text{off},i}$ do not depend on P. The burst frequencies can be represented by sigmoid functions [23] as a simplification of the mechanistic form used in [6, 14]:

$$k^{\theta}_{\text{on},i}(P) = k_{0,i} + (k_{1,i} - k_{0,i}) \left(1 + \exp\left(-\beta_i - \sum_{j=1}^{n} \theta_{ij} P_j\right)\right)^{-1}$$

where $k_{0,i}$ (resp. $k_{1,i}$) is the minimal (resp. maximal) burst frequency of gene i and β_i is the basal activity of gene i, which can be also considered as the constant activity of a set of genes that are not measured but act on the network.

Simulation of time-stamped datasets

In order to simulate the mechanistic model, we used the simulation module of the HARISSA package [23]. One computational advantage of this method, which consists in sampling burst times with maximum rate and then deciding with an appropriate rule which ones to keep, is that it is guaranteed to be exact without requiring any numerical integration.

To simulate discrete "count" data that are produced by current scRNA-seq technologies, each mRNA level is generated by sampling from a Poisson distribution whose mean is the simulated expression level. The resulting cell profiles are exactly (resp. approximately) distributed according

to the discrete-valued "Gillespie" version of the mechanistic model in the absence (presence) of interactions between genes [52].

In order to reproduce *in vitro* experiments for a specific GRN, we use the following method: (1) let the model run for t $<$ 0h until its (stochastic) steady state is reached; (2) introduce at t = 0h a virtual *stimulus* gene with a constant maximal value for its protein. Such stimulus represents a perturbation in the environment of the cells, inducing them to evolve towards a new (stochastic) steady state. For example, in the case of mouse ES cell differentiation, this corresponds to the addition of all-trans RA in the medium. A time-stamped dataset corresponds to the sampling of independent cells at a specific sequence of timepoints (therefore "killing" sampled cells at each timepoint) starting from t = 0h. Namely, for the benchmark of Figure 2, the sequence of 10 timepoints was set to 0, 6, 12, 24, 36, 48, 60, 72, 84, and 96h.

Relevance to biological data

The exact probability distribution associated to the mechanistic model remains unknown for general networks. However, the analysis developed in [15] suggests that the marginal on mRNAs of the distribution at each time t can be reasonably approximated by a Gamma mixture:

$$M_t \sim \sum_{z \in Z} \mu_t(z) \prod_{i=1}^{n} \text{Gamma}\left(\frac{k_{z,i}}{d_{0,i}}, \frac{k_{\text{off},i}}{s_{0,i}}\right), \tag{1}$$

where Z denotes the set of cell types seen as the basins of attraction of the deterministic limit, $k_{z,j}$ denotes the mode of bursts frequency associated to a gene j within a basin $z \in Z$, and μ_t is a probability vector describing the relative weight of the basins in the process at time t.

The Poissonian layer transforms the Gamma distributions into negative binomial (NB) distributions, which gives:

$$M_t \sim \sum_{z \in Z} \mu_t(z) \prod_{i=1}^{n} \text{NB}\left(\frac{k_{z,i}}{d_{0,i}}, \frac{k_{\text{off},i}}{k_{\text{off},i} + s_{0,i}}\right). \tag{2}$$

Such mixture distributions are known to be compatible with continuous single-cell data [1, 10]. In particular, we recover the second order relationship between pairs of variables that are characteristic of experimental datasets. Indeed, we remark that the mean of a negative binomial distribution NB(a, b) is $m = \frac{a(1-b)}{b}$ and its variance $v = \frac{a(1-b)}{b^2}$, which implies that:

$$CV^2 = \frac{v}{m^2} = \frac{1}{a(1-b)} = \frac{1}{b}\frac{1}{m}.$$

Thus, for every gene i, by replacing b by $\frac{k_{\text{off},i}}{k_{\text{off},i}+s_{0,i}}$, we see that the relation $CV^2 \sim \frac{1}{m}$, which is characteristic of cell-cell heterogeneity in single-cell data is well verified by the mechanistic model, provided that $k_{\text{off},i}$ does not depend on the protein field. This also argues in favor of the assumption that a GRN does not affect significantly the bursts size. Note however that following this criteria, any model generating negative binomial distributions could be considered as realistic. Such criteria is therefore not sufficient for characterizing the accuracy of a model of gene expression, in particular when a synthetic noise well adapted is added to be in accordance with experimental datasets [16].

Tested algorithms

The six algorithms that we are going to use for the benchmark represent together the main categories of GRN inference methods presented in [7]:

- GENIE3 [18], which computes the regulatory network for each gene independently, using tree-based ensemble methods to predict the expression profile of each target gene from the profiles of all the other genes;
- PIDC [19], which infers undirected network using the notion of mutual information;
- SINCERITIES [20], which uses Granger causality after computing temporal changes in gene expression through the distance between two consecutive timepoints of the marginal distributions;
- SCRIBE [21], which is based on the notion of conditioned Restricted Directed Information and ideally needs real cells trajectories, which is unrealistic experimentally. We then pseudo-temporally order the time-stamped synthetic data used for the benchmark with two methods, one using a pseudotime algorithm and the other using an optimal coupling method with optimal transport, following the idea developed in [27]. We also tested with a dataset with real trajectories in order to compare the performances. The results of this algorithm are presented separately, due to the difference in the information that is needed.
- HARISSA [15] and CARDAMOM [23] which are based on the mechanistic model presented above, and both compute the network by solving a set of regression problems. They are nevertheless based on distinct mathematical analyses of the same model. HARISSA aims to solve a maximum likelihood problem on the protein distributions after computing a most-probable position for the protein concentration in each cell, CARDAMOM reconstructs the GRN by comparing the function k_{on} to the modes associated to a joint mRNA distribution previously inferred.

Measuring algorithm performance for the benchmark

We evaluated the GRN inference algorithms on simulated datasets using the area under the precision-recall curve (AUPR). Since inferring these coefficients is a notoriously difficult task [17], we do not take into account diagonal coefficients of the GRN matrix, which correspond to self-regulations. Note also that we chose precision-recall (PR) curves rather than receiver operating characteristic (ROC) curves because of the well-known class imbalance problem. Indeed, the sparsity hypothesis suggests that the number of interactions expected for a network of size n is smaller than half of the total number of possible interactions (n^2): it is then natural to focus on minimizing false positives (interactions that are detected but not present) rather than false negatives (interactions that are present but not detected), which explains the preference of PR over ROC.

Experimental dataset

We used data collected from a differentiation experiment of mouse embryonic stem cells induced by all-trans retinoic acid treatment [22]. This scRNA-seq dataset consists of 9 timepoints (0,

6, 12, 24, 36, 48, 60, 72, and 96h), each timepoint containing between 137 and 335 sampled cells after pre-processing (272 on average, for a total number of 2449 cells). To limit artificial correlations between genes (due to a multiplicative cell-specific technical factor mainly related to the reverse-transcription step), we selected cells with a total number of UMI counts ≥ 2000 in line with Semrau et al. [22], which resulted in keeping only 2449 out of 3456 measured cells.

On the other hand, we did not normalize cells by their respective total UMI counts and argue that this type of sample normalization is hazardous in the case of single-cell data. Indeed, such "library sizes" are small compared to bulk data (because here 1 sample = 1 cell) and are in fact biologically fluctuating, likely reflecting the transcriptional bursting phenomenon (this is easily seen when simulating "perfect" data from the mechanistic model, see Figure S2). In practice, since the CARDAMOM inference method starts with a binarization step (applying a specific, statistically derived threshold to each gene based on the mechanistic model), a multiplicative factor on each cell should not have too much impact as long as the number of cells is large enough. More generally, we argue that such normalization of cells should rather be "soft-coded" as a random factor to be estimated within a statistical framework.

The total number of genes measured in this experiment is 17452, which is much larger than in our benchmark. As they are unlikely to all be important in characterizing the differentiation process, we decided to restrict our analysis to a panel of 41 genes that had previously been identified as key marker genes for pluripotency, post-implantation epiblast, neuroectoderm and extraembryonic endoderm [22]. This number of genes allows to infer a network rich enough to make cell types emerge in a non-trivial way, while keeping a reasonable statistical power regarding the number of sampled cells. Note that the speed of the algorithms (Table S1) would allow a much larger subset of genes to be used: the limiting factor here is not computational speed but statistical power resulting from the number of cells available (see Figure 2).

Calibration of the mechanistic model

The principle of CARDAMOM is based on a two-step procedure :

- In a first step, we find the set of parameters α defining the mixture of negative binomial distributions (2) which fits well the data. The only parameter which is allowed to vary at each timepoint is the mixture parameters (allowing to estimate the values of the typical modes associated to the functions $k_{\text{on},i}$, for every gene i). Note that every parameters are computed except the degradation rates, which are constant for each gene and scale the dynamics of gene expression;

- In a second step, we calibrate the mechanistic model in order to obtain the distribution which is the closest to this Gamma mixture distribution. The network θ is then actualized at each timepoint in order to fit the mixture parameters.

The degradation rates are then fixed at the values that are found in tables from the literature [36]. Since many genes do not appear in these tables, we decide to set the same value for the degradation rates of all the genes belonging to the same functional group identified in [22] (see Table S2). The detail concerning these two steps can be found in [15]. Note that the first step has been modified since the original publication in order to replace the MCMC algorithm, which

was used to find the parameters of the negative binomial distributions associated to each cell type, by a simpler variational method.

As mentioned previously, the model cannot be in adequacy with both the dynamics from 0h to 72h and from 72h to 96h with the same degradation rates for the experimental dataset used in Results, due to the acceleration of the process between 72h and 96h. For this reason, we decided for our simulation to multiply degradation rates before the last timepoint by a factor $f = 6$, large enough to reach the steady state between 72h and 96h.

Comparing datasets using countsimQC

We used countsimQC [37] to compare the experimental dataset, the dataset simulated with the inferred network and the dataset simulated with the null network (Figure S2). Although there are no significant difference between the first two datasets, we observe in Figure S2F, that the correlations between genes are not perfectly reproduced (they are clearly more accurate than for the dataset simulated with the null network). This gap between the correlations between genes is also illustrated in Figure S4, which compares joint distributions between the simulated dataset (with the inferred network) and the experimental one for three pairs of genes at the final timepoint. We observe that if the global form of the correlation is respected, they are not as strong in the simulated dataset as in the experimental dataset. This suggests that the inferred GRN recovers the true correlation but not with the right intensity, which may be due to the sensitivity of model to the value of its parameters.

The fact that except for the correlations (sample-sample and feature-feature), the statistical characteristics explored by countsimQC are also well reproduced by the dataset simulated with the null network, suggests that they are generally not sufficient for measuring the accuracy of a dataset reproduction. This is partly due to the fact that any calibration of the model with the right scaling parameters but not the right GRN should matches most of these characteristics. In that meaning, the successes of simulation algorithms prior to our work are limited when measured with similar criteria. Our methodology, for which we used distinct criteria which are particularly well illustrated in Figure S5 and Figure 7, then appears as a significant improvement in the field of executable GRN inference.

Comparing datasets using UMAP

Since UMAP is not linear, the projections of datasets shown in Figure S5 are likely to force the projected data to be artificially close to the reference dataset. Thus, we decide to present two figures similar to Figure 7E-F-G, but where instead of projecting the simulated dataset on the pseudo-axis corresponding to the projection of the experimental dataset, we project both datasets together and show separately the cells corresponding to each dataset. Using this methodology, we represented in Figure S5A the UMAP projection of the experimental dataset and in Figure S5B the one of the dataset simulated with the inferred network. We did the same for the experimental dataset (Figure S5C) and the dataset simulated with the null network (Figure S5D). Then, although they have different representations, Figures Figure S5A and Figure S5C represent the same dataset, and the difference comes from the second dataset (the simulated one) with which the reduction has been performed. This allows to emphasize that the representation of a

distribution of cells with UMAP is very sensitive to the choice of the data that are integrated in the projection. Once again, we observed that the dataset simulated using the inferred network does seem much closer to the experimental dataset than the one simulated with the null network: in particular, Figure S5B demonstrates that the UMAP projection of the dataset with network is close to the one of the experimental dataset both in the arrangement of the cells between the different timepoints and in the general form of the subspace occupied by the cells, which is not the case for the UMAP projection of the dataset simulated with the null network, represented in Figure S5D.

Code availability

CARDAMOM is available at `https://gitbio.ens-lyon.fr/eventr01/cardamom`.
HARISSA is available at `https://github.com/ulysseherbach/harissa`.

Author contributions

Conceptualization, E.V., U.H., T.E., and O.G.; Formal Analysis, E.V., U.H., and G.B.; Funding Acquisition, O.G.; Investigation, All; Methodology, E.V., U.H., and T.E.; Software, E.V. and U.H.; Supervision, O.G. and T.E.; Validation, All; Visualization, E.V., U.H., and O.G.; Writing – Original Draft, E.V., G.B., and O.G.; Writing – Review and Editing, All.

Acknowledgments

This work was supported by funding from French agency ANR (SingleStatOmics; ANR-18-CE45-0023-03). We would like to thank especially Christophe Arpin, Thomas Lepoutre, Anton Crombach and Arnaud Bonnafoux for critical reading of the manuscript. We also thank all members of the SBDM and Dracula teams, and of the SingleStatOmics project, for providing such stimulating working environment. We finally thank the BioSyL Federation, the LabEx Ecofect (ANR-11-LABX-0048) and the LabEx Milyon of the University of Lyon for inspiring scientific events.

References

[1] J. C. Mar. “The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond”. In: *Biophysical Reviews* 11 (2019), pp. 89–94.

[2] A. F. Coskun, U. Eser, and S. Islam. “Cellular identity at the single-cell level”. In: *Mol Biosyst* 12 (2016), pp. 2965–2979.

[3] V. A. Huynh-Thu and G. Sanguinetti. “Gene Regulatory Network Inference: An Introductory Survey”. In: *Methods Mol Biol* 1883 (2019), pp. 1–23.

[4] C. Zong, L.-H. So, L. A Sepúlveda, S. O Skinner, and I. Golding. “Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene”. In: *Molecular systems biology* 6 (2010), p. 440.

[5] H. Ochiai, T. Sugawara, T. Sakuma, and T. Yamamoto. “Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells”. In: *Scientific reports* 4 (2014), pp. 1–9.

[6] U. Herbach, A. Bonnaffoux, T. Espinasse, and O. Gandrillon. “Inferring gene regulatory networks from single-cell data: a mechanistic approach”. In: *BMC Systems Biology* 11 (2017), p. 105.

[7] K. Akers and T.M. Murali. “Gene regulatory network inference in single cell biology”. In: *Current Opinion in Systems Biology* (2021).

[8] V. Shahrezaei and P. S. Swain. “Analytical distributions for stochastic gene expression”. In: *PNAS* 105 (2008), pp. 17256–17261.

[9] N. Friedman, L. Cai, and X S. Xie. “Linking stochastic dynamics to population distribution: an analytical framework of gene expression”. In: *Phys Rev Lett* 97 (2006).

[10] C. Albayrak, C. A. Jordi, C. Zechner, J. Lin, C. A. Bichsel, M. Khammash, and S. Tay. “Digital Quantification of Proteins and mRNA in Single Mammalian Cells”. In: *Molecular Cell* 61 (2016), pp. 914–924.

[11] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz. “Dynamic heterogeneity and DNA methylation in embryonic stem cells”. In: *Mol Cell* 55 (2014), pp. 319–31.

[12] G. Schiebinger. “Reconstructing developmental landscapes and trajectories from single-cell data”. In: *Current Opinion in Systems Biology* 27 (2021), p. 100351.

[13] L. Deconinck, R. Cannoodt, W. Saelens, B. Deplancke, and Y. Saeys. “Recent advances in trajectory inference from single-cell omics data”. In: *Current Opinion in Systems Biology* 27 (2021), p. 100344.

[14] A. Bonnaffoux, U. Herbach, A. Richard, A. Guillemin, S. Gonin-Giraud, P.-A. Gros, and O. Gandrillon. “WASABI: a dynamic iterative framework for gene regulatory network inference”. In: *BMC Bioinformatics* 20 (2019), p. 220.

[15] E. Ventre. “Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics”. In: *In Silico Biology* 14 (2021), pp. 89–113.

[16] P. Dibaeinia and S. Sinha. “SERGIO: A Single-Cell Expression Simulator Guided by Gene Regulatory Networks”. In: *Cell systems* (2020).

[17] A. Pratapa, A. P. Jalihal, J. N. Law, A. Bharadwaj, and T. M. Murali. “Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data”. In: *Nat Methods* 17 (2020), pp. 147–154.

[18] V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts. “Inferring regulatory networks from expression data using tree-based methods”. In: *PLOS One* 5 (2010).

[19] T. E Chan, M. P H Stumpf, and A. C Babtie. “Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures”. In: *Cell Systems* 5 (2017).

[20] N. Papili Gao, S. M. M. Ud-Dean, O. Gandrillon, and R. Gunawan. “SINCERITIES: Inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles”. In: *Bioinformatics* (2017).

[21] X. Qiu, A. Rahimzamani, L. Wang, B. Ren, Q. Mao, T. Durham, J.L. McFaline-Figueroa, L. Saunders, C. Trapnell, and S. Kannan. “Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe”. In: *Cell Systems* 10 (2020), pp. 1–10.

[22] S. Semrau, J. E. Goldmann, M. Soumillon, T. S. Mikkelsen, R. Jaenisch, and A. van Oudenaarden. “Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells”. In: *Nat Commun* 8 (2017), pp. 1–16.

[23] U. Herbach. “Gene regulatory network inference from single-cell data using a self-consistent proteomic field”. In: *arXiv* 2109.14888 (2021).

[24] A. Richard, L. Boullu, U. Herbach, A. Bonnafoux, V. Morin, E. Vallin, A. Guillemin, N. Papili Gao, R. Gunawan, J. Cosette, O. Arnaud, J. J. Kupiec, T. Espinasse, S. Gonin-Giraud, and O. Gandrillon. "Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process". In: *PLoS Biol* 14 (2016), e1002585.

[25] P.S. Stumpf, R.C.G. Smith, M. Lenz, A. Schuppert, F.-J. Müller, A Babtie, T.E. Chan, M.P.H. Stumpf, C.P. Please, S.D. Howison, F. Arai, and B.D. MacArthur. "Stem Cell Differentiation as a Non-Markov Stochastic Process". In: *Cell Systems* 5 (2017), pp. 268–282.

[26] N. E. Phillips, A. Mandic, S. Omidi, F. Naef, and D. M. Suter. "Memory and relatedness of transcriptional activity in mammalian cell lineages". In: *Nature Communications* 10 (2019).

[27] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, and E. S. Lander. "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming". In: *Cell* 176 (2019), 928–943 e22.

[28] K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and S. Dudoit. "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics". In: *BMC Genomics* 19 (2018), p. 477.

[29] E. Moutier, T. Ye, M. A. Choukrallah, S. Urban, J. Osz, A. Chatagnon, L. Delacroix, D. Langer, N. Rochel, D. Moras, G. Benoit, and I. Davidson. "Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology". In: *J Biol Chem* 287 (2012), pp. 26328–41.

[30] A. Chatagnon, P. Veber, V. Morin, J. Bedo, G. Triqueneaux, M. Semon, V. Laudet, F. d'Alche-Buc, and G. Benoit. "RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements". In: *Nucleic Acids Res* (2015).

[31] E. H. Z. Chua, S. Yasar, and N. Harmston. "The importance of considering regulatory domains in genome-wide analyses - the nearest gene is often wrong!" In: *Biol Open* 11.4 (2022).

[32] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng. "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells". In: *Cell* 133 (2008), pp. 1106–17.

[33] G. Li, R. Margueron, M. Ku, P. Chambon, B. E Bernstein, and D. Reinberg. "Jarid2 and PRC2, partners in regulating gene expression". In: *Genes & development* 24 (2010), pp. 368–380.

[34] S. Mahony, E. O Mazzoni, S. McCuine, R. A Young, H. Wichterle, and D. K Gifford. "Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis". In: *Genome biology* 12 (2011), pp. 1–15.

[35] C. Hrabchak, M. Ringuette, and K. Woodhouse. "Recombinant mouse SPARC promotes parietal endoderm differentiation and cardiomyogenesis in embryoid bodies". In: *Biochemistry and Cell Biology* 86 (2008), pp. 487–499.

[36] B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. "Global quantification of mammalian gene expression control". In: *Nature* 473 (2011), pp. 337–42.

[37] C. Soneson and M. D Robinson. "Towards unified quality verification of synthetic count data with countsimQC". In: *Bioinformatics* 34 (2018), pp. 691–692.

[38] H. L. Crowell, S. X. M. Leonardo, C. Soneson, and M. D. Robinson. "Built on sand: the shaky foundations of simulating single-cell RNA sequencing data". In: *bioRxiv* (2021).

[39] K. S. Manning and T. A. Cooper. "The roles of RNA processing in translating genotype to phenotype". In: *Nat Rev Mol Cell Biol* 18 (2017), pp. 102–114.

[40] L. McInnes, J. Healy, and J. Melville. “UMAP: uniform manifold approximation and projection for dimension reduction”. In: *arXiv* 1802.03426 (2020).

[41] R. Cannoodt, W. Saelens, L. Deconinck, and Y. Saeys. “Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells”. In: *Nature Communications* 12.1 (2021).

[42] A. Mizeranschi, H. Zheng, P. Thompson, and W. Dubitzky. “Evaluating a common semi-mechanistic mathematical model of gene-regulatory networks”. In: *BMC Systems Biology* 9 (2015), pp. 1–12.

[43] H. Matsumoto, H. Kiryu, C. Furusawa, M. S. Ko, S. B. Ko, N. Gouda, T. Hayashi, and I. Nikaido. “SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation”. In: *Bioinformatics* (2017).

[44] E. Ventre, T. Espinasse, C.-E. Bréhier, V. Calvez, T. Lepoutre, and O. Gandrillon. “Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastabilty”. In: *Journal of Mathematical Biology* 83 (2021), pp. 1–63.

[45] M. Marouf, P. Machart, V. Bansal, C. Kilian, D. S. Magruder, C. F. Krebs, and S. Bonn. “Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks”. In: *Nat Commun* 11 (2020), p. 166.

[46] H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger. “Towards a mathematical theory of trajectory inference”. In: *arXiv preprint arXiv:2102.09204* (2021).

[47] D. M Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. “Mammalian genes are transcribed with widely different bursting kinetics”. In: *Science* 332 (2011).

[48] D. Nicolas, N. E. Phillips, and F. Naef. “What shapes eukaryotic transcriptional bursting?” In: *Mol Biosyst* 13 (2017), pp. 1280–1290.

[49] J. Rodriguez and D. R. Larson. “Transcription in Living Cells: Molecular Mechanisms of Bursting”. In: *Annu Rev Biochem* 89 (2020), pp. 189–212.

[50] E. Tunnacliffe and J. R. Chubb. “What Is a Transcriptional Burst?” In: *Trends Genet* 36 (2020), pp. 288–297.

[51] C. Li, F. Cesbron, M. Oehler, M. Brunner, and T. Hofer. “Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation”. In: *Cell Syst* (2018).

[52] U. Herbach. “Stochastic gene expression with a multistate promoter: breaking down exact distributions”. In: *SIAM Journal on Applied Mathematics* 79 (2019), pp. 1007–1029.

Supplementary information

Table S1: Related to Figure 2. Average runtime for inferring a network from datasets simulated with tree-like networks for which the results of the inference are represented in Figure 2, for the six algorithms that are used in the benchmark. Timings measured on a 16-GB RAM, 2.4 GHz Intel Core i5 computer.

Runtime	PIDC	SINCERITIES	HARISSA	CARDAMOM	GENIE3	SCRIBE
5 genes	0.02 s	0.01 s	0.16 s	0.24 s	9.02 s	34.36 s
10 genes	0.04 s	0.03 s	0.28 s	0.64 s	17.27 s	136.51 s
20 genes	0.06 s	0.06 s	0.45 s	1.70 s	31.17 s	> 5 min
50 genes	0.18 s	0.24 s	0.96 s	7.87 s	83.18 s	> 1 h
100 genes	0.75 s	0.77 s	1.80 s	18.57 s	159.42 s	> 3 h

Table S2: Related to Figure 7. Numerical values of mRNA and protein degradation rates (in h^{-1}) used for data simulation (sim.), compared with experimental (exp.) measures from the literature. Group abbreviations: Pluri = Pluripotency, Epi = Post-implantation epiblast, Neuro = Neuroectoderm, Endo = Extraembryonic endoderm.

Gene	Sox2	Zfp42	Klf2	Dnmt3	Cdh2	Sparc	Col4a2	Lamb1
Group	Pluri			Epi	Neuro	Endo		
mRNA (exp.)	0.044	0.053	0.067	-	0.015	0.012	0.012	0.015
mRNA (sim.)	0.075			0.2	0.03	0.005		
Protein (exp.)	0.0073	0.05	0.0027	0.023	0.022	0.17	0.078	0.038
Protein (sim.)	0.0075			0.02	0.02	0.1		

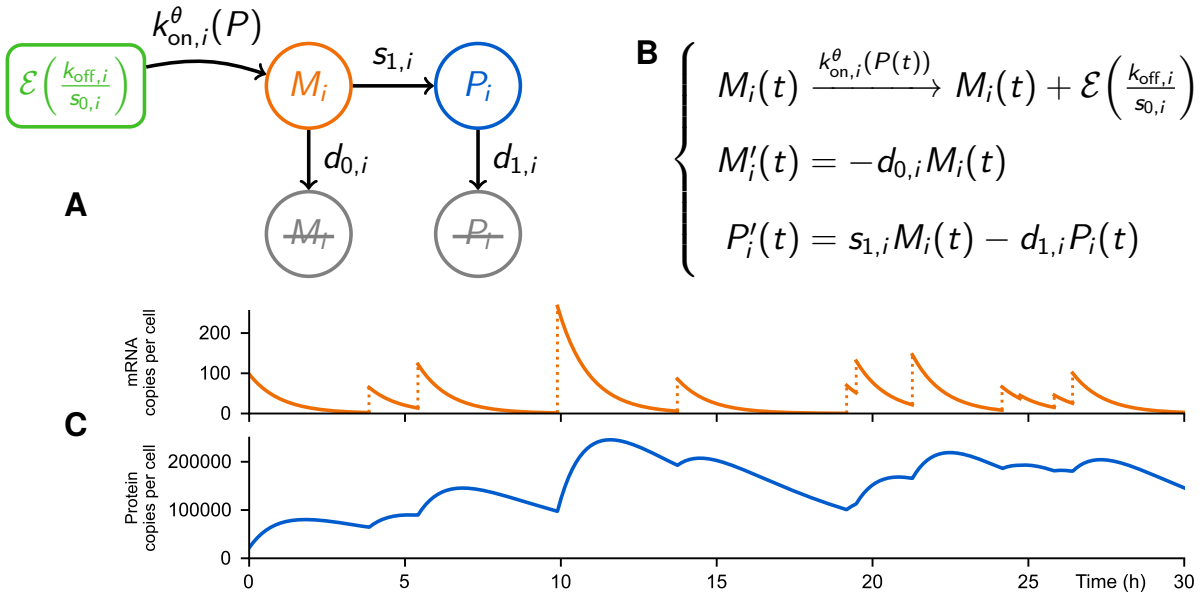


Figure S1: Related to Figure 1. Graphical (**A**) and mathematical (**B**) descriptions of the mechanistic model for the dynamics of a gene i. (**C**) Bursts of mRNA occur at random times with rate $k_{\text{on},i}$ and their size follows an exponential distribution $\mathcal{E}(k_{\text{off},i}/s_{0,i})$. The variables M_i and P_i describe respectively the mRNA and protein quantities associated to gene i in the cell. The vector of protein levels is denoted by $P = (P_1, \cdots, P_n)$ while θ denotes the GRN which couples the genes together through functions $k_{\text{on},i}$.

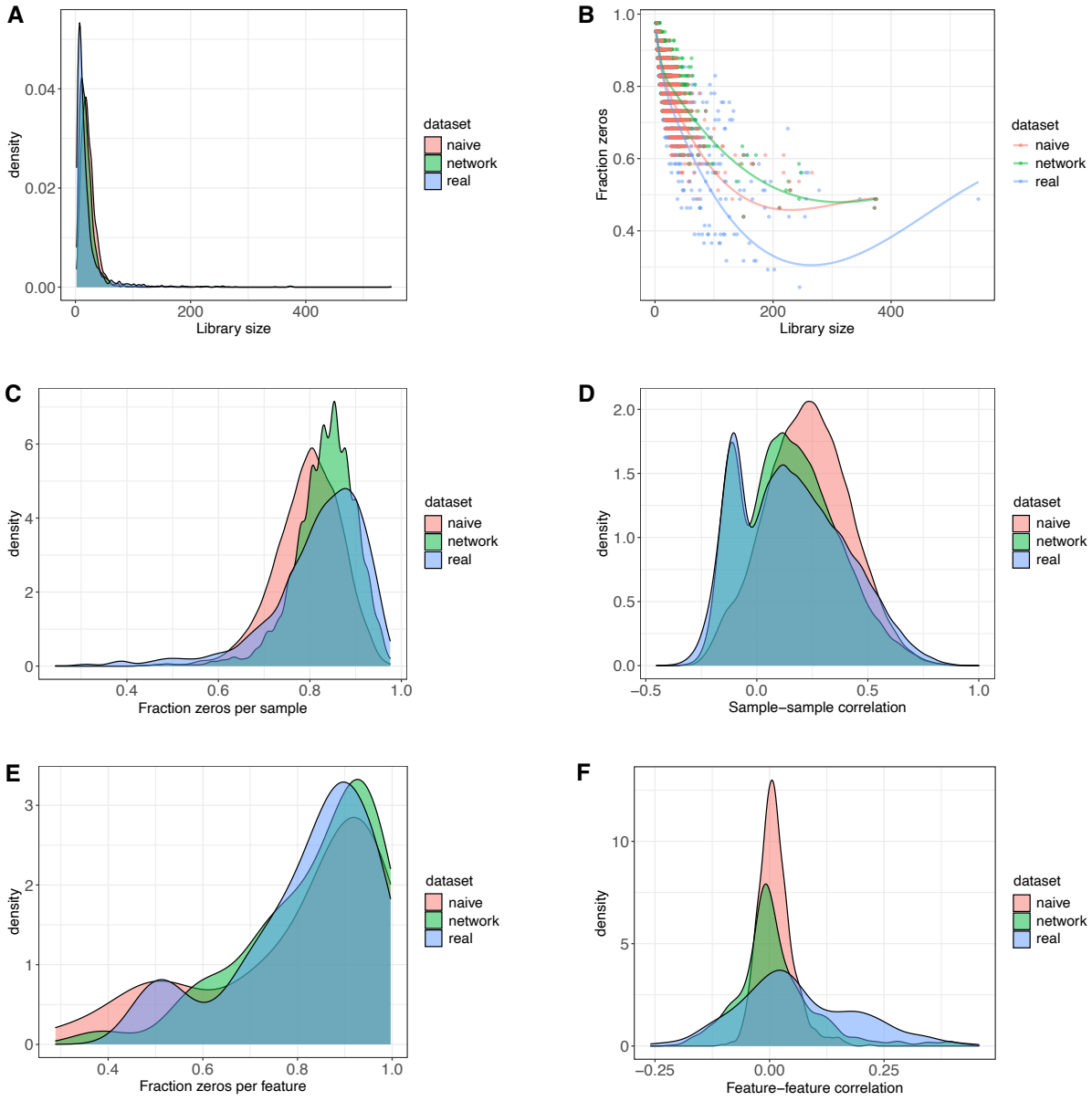


Figure S2: Related to Figure 7. Comparison of various statistical characteristics across datasets using the countsimQC package. Each plot shows the experimental dataset (*real*, in blue) and datasets simulated from the mechanistic model calibrated by CARDAMOM, including interactions (*network*, in green) and without interactions (*naive*, in red). Each dataset consists of 41 genes (*features*) measured in 2433 single cells (*samples*). (**A**) Distribution of "library sizes", defined as the total read count in each sample. (**B**) Association between the library size and the fraction of zeros observed per cell. (**C**) Distribution of the fraction of zeros observed per cell. (**D**) Distribution of cell-cell correlations, based on random cell pairings. (**E**) Distribution of the fraction of zeros observed per gene. (**F**) Distribution of gene-gene correlations, based on all possible gene pairings. Notably, the cell-cell correlation (**D**) bimodal pattern shows two possible pairings of cells: pairs with similar expression profiles (same genes *on*, same genes *off*) and therefore positively correlated, and pairs with opposite, "antinomic" profiles and therefore negatively correlated. This pattern is an indirect sign of the emergence of different cell types, a characteristic that is clearly not reproduced in the absence of interactions between genes (naive dataset).

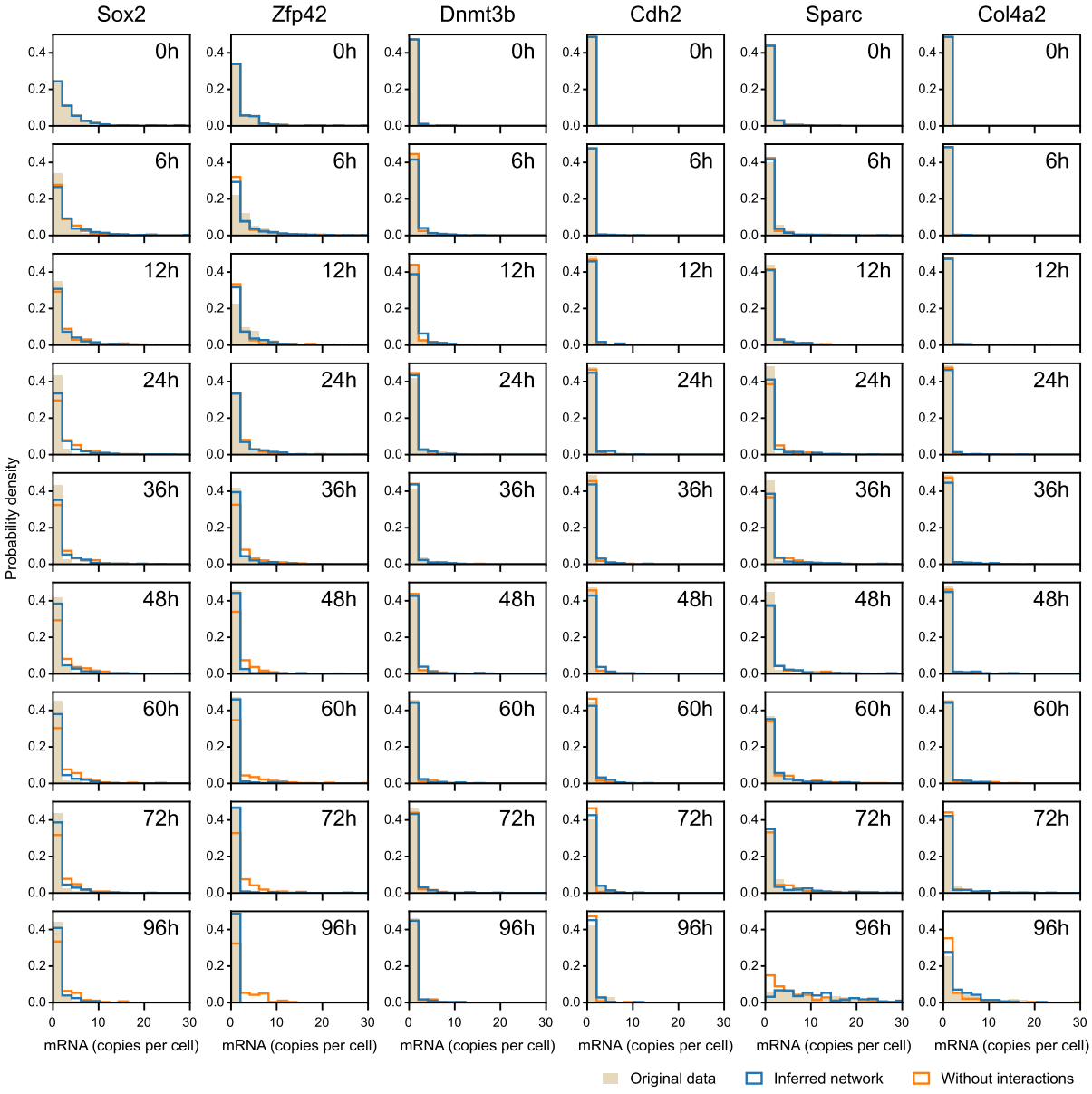


Figure S3: Related to Figure 7. Comparison between empirical distributions along timepoints, for six genes that have been found to play a key role in the regulation of the process (as visible in Figure 5). The experimental dataset (in beige), the dataset simulated from the inferred network (in blue) and the dataset simulated without interactions (in orange) correspond to Figure 7E, F and G, respectively.

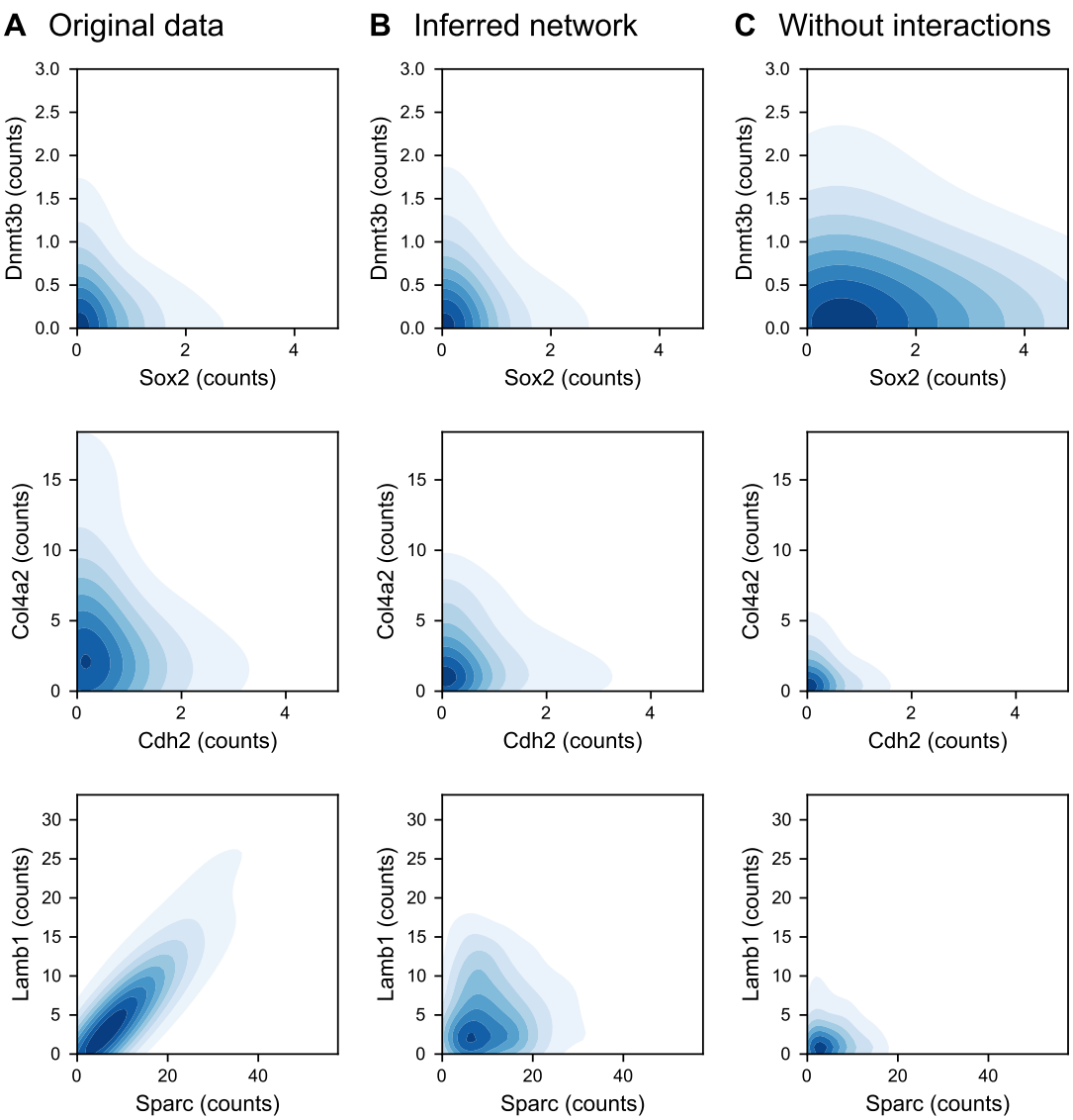


Figure S4: Related to Figure 7. Comparison of the joint distributions of three pairs of genes at the final timepoint between the experimental dataset (**A**) compared to the dataset simulated when the mechanistic model is calibrated by CARDAMOM (**B**) and the dataset simulated without interactions (**C**). The genes in each pair are expected to have interactions (direct or indirect) in the network represented in Figure 5.

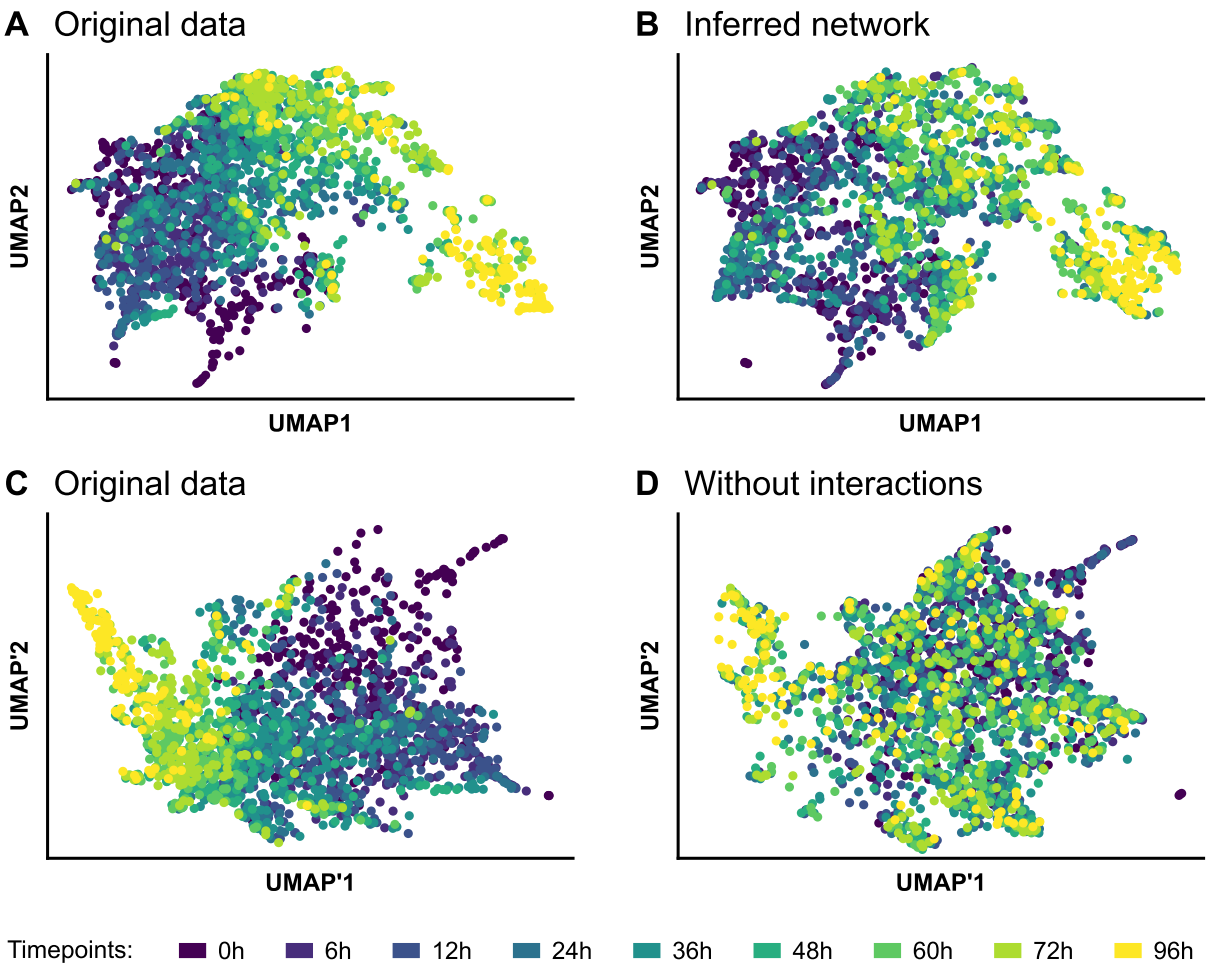


Figure S5: Related to Figure 7. Two-dimensional UMAP representations of the experimental dataset (*original data*) and datasets simulated after calibrating the mechanistic model with CARDAMOM, one including interactions (*inferred network*) and one obtained after setting $\theta = 0$ (*without interactions*). The four plots are based on two different projections, computed after merging the experimental dataset with (**A-B**) the simulated dataset including interactions or (**C-D**) the simulated dataset without interactions. Hence **A** and **C** are two representations of the exact same data, while **B** is to be compared with **A**, and **D** is to be compared with **C**.

Part III

A mechanistic approach of entropy minimization problems for single-cell gene expression analyzes.

We have seen in Part II that it was possible, from time-stamped datasets, to efficiently reverse-engineer the bursty model in order to simulate synthetic datasets that accurately reproduce the original data. In Chapter 4, the accuracy of this reproduction was measured using different criteria, including the p-value of a statistical test between the experimental and simulated empirical distributions on the marginals of each gene, the comparison of UMAP projections and Wasserstein distances between the datasets. However, it remains unclear whether these metrics are adapted or not for these data (especially for UMAP and Wasserstein distance, which may appear as arbitrary choices) and it is then difficult to precisely quantify the differences between simulations and observations. Moreover, the CARDAMOM algorithm that we developed for reverse-engineering the data, although it seems very efficient in practice, is not exact in the sense that it does not give any theoretical guarantee of convergence to an "optimal" calibration when the number of observed cells is infinite. Thus, it would be of great interest not only to quantify the differences between observations and simulations but also to answer key questions like:

1. Is the model well calibrated with respect to the data ?
2. If the calibration is not optimal with respect to data, how to find the optimal parameters ?
3. Is the model good enough for reproducing the data with an optimal calibration ?

Remark that the last question is more subtle, as it implies to distinguish an error due to the calibration from an error due to the limitations of the model, which could be hard when the calibration problem is not clearly identifiable.

In this last part, we propose to use the notion of relative entropy for addressing these issues. As mentioned in the introduction, the choice of the relative entropy is very natural for analyzing stochastic processes (Sections 1.1.2 and 1.1.2), and has already been used in pioneer works, as well as its connections with the theory of optimal transport [84, 47]. However, although inspired by these studies, we are going to use a slightly different approach. We are not going to consider a Brownian motion as the reference process, but rather consider a realistic bursty process of the form (1.17), and try to address the issues 1-2-3. In that context, the parameters of the reference process can be seen as a prior knowledge on the system. Typically, one may consider that the process is calibrated by CARDAMOM in a first step, and the method then consists in solving a Schrödinger problem with this reference. When the Schrödinger problem has a solution, we know that the associated dynamical formulation (1.5) has a solution too, that is related to the one of the Schrödinger problem by Theorem 7, and we obtain a cost characterizing the distance of the reference process to the observations.
However, since the bursty model is constrained by dynamical parameters, the data in the Schrödinger problem may be such that the latter admits no solution. We then provide a general point of view in Chapter 5, showing that whatever be the reference process and the observations, the most classical algorithm for solving Schrödinger problems –the Sinkhorn algorithm– converges to exactly two limit points, each of them being the solution of a problem with modified constraints, that characterize themselves relevant auxiliary optimization problems.
In Chapter 6, we analyze the link between the solution of the Schrödinger problem and the associated dynamical problem, in order to deduce the optimal modifications of the reference process with respect to the observations. In particular, we show how to find the optimal jump kernel associated to any couple of experimental observations and a reference bursty process, using the results of Chapter 5. We also propose a method for reconstructing the optimal parameters, and in particular a GRN, associated to this optimal kernel. Note that our results suggest that this method seems to be nevertheless more adapted for *evaluating* the accuracy of a model which would have been yet calibrated than for *calibrating* the model directly.

Chapter 5

Resolution of the Schrödinger problem when it has no solution. Preprint available on arXiv.

The Schrödinger problem (1.4) is going to be the cornerstone of the approach developed in the next section for assessing the model calibration against gene expression datasets. We place this chapter in a discrete setting, and consider that the reference joint distribution is simply a matrix $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mathcal{D}$ and $\mathcal{F}$ being to finite subsets of the gene expression space. In that case, according to the Sanov theorem, when the number of observations is large enough, the solution of the Schrödinger problem provides an estimation of the coupling of the process knowing to the observations.
It is known that if the reference matrix R has nonnegative but possibly cancelling entries, observations could happen to be such that there exists no coupling compatible with the observed marginals while being absolutely continuous w.r.t R (and then of finite entropy): the Schrödinger problem would thus have no solution. This happens to be precisely the case for the mechanistic models of gene expression like (1.17): indeed, the dynamical constraints involving mRNAs half-life times can lead to a degenerate R such that the corresponding Schrödinger problem has no solution, not because of a lack in the model but because of inaccuracies in the measurements. We would like to characterize nevertheless these small measurement errors and be able to find a coupling which explains the best the data while being absolutely continuous with respect to R. Although alternative methods are available in such cases, in particular an algorithm which solves the so-called unbalanced problem [17], it is not totally satisfying for our purposes. In addition to the fact that it introduces a new parameter quantifying the balance between the proximity to the data and to the reference coupling, whose value will often be arbitrarily chosen, the relation with the corresponding dynamical Schrödinger problem is not clear.
We show in this article that interestingly, the most popular algorithm for solving the Schrödinger problem, the Sinkhorn algorithm, is still adapted when the problem has no solution. Our main finding is that it leads to exactly two limit points, each of them being the solution of a Schrödinger problem with modified data, that we characterize as solutions of auxiliary optimization problems. Also, we show that these limit points are related to a problem where the marginal constraints of the original problem are replaced by marginal penalizations. We therefore provide a new outlook on the question of the support of the solution in this case, allowing to design an approximate method for improving the Sinkhorn algorithm's convergence in cases where it is not linear.

This chapter contains a preprint which is available on arXiv [8].

CONVERGENCE OF THE SINKHORN ALGORITHM WHEN THE SCHRÖDINGER PROBLEM HAS NO SOLUTION

Authors: Aymeric Baradat[1], Elias Ventre[1,2,3].
1 - Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, Villeurbanne, France.
2 - ENS de Lyon, CNRS UMR 5239, Laboratory of Biology and Modelling of the Cell, Lyon, France.
3 - Inria Center Grenoble Rhone-Alpes, Equipe Dracula, Villeurbanne, France.
Corresponding author: aymeric.baradat@cnrs.fr.

Abstract

The Sinkhorn algorithm is the most popular method for solving the Schrödinger problem: it is known to converge as soon as the latter has a solution, and with a linear rate when the solution has the same support as the reference coupling. Motivated by recent applications of the Schrödinger problem where structured stochastic processes lead to degenerate situations with possibly no solution, we show that the Sinkhorn algorithm still gives rise in this case to exactly two limit points, that can be used to compute the solution of a relaxed version of the Schrödinger problem, which appears as the Γ-limit of a problem where the marginal constraints are replaced by marginal penalizations. These results also allow to develop a theoretical procedure for characterizing the support of the solution – both in the original and in the relaxed problem – for any reference coupling and marginal constraints. We showcase promising numerical applications related to a model used in cell biology.

Keywords: Schrödinger problem, the Sinkhorn algorithm, matrix scaling

Contents

1 Introduction

The Schrödinger problem has been introduced by Schrödinger himself in the 30's [28, 27] in the context of statistical mechanics. It is one of these problems in mathematics for which there is periodically a resurgence of interest, as witnessed by the numerous works which it was the object of for almost 100 years, among which [10, 7, 34, 9, 24]. The last of these resurgences, over the past twenty years, occurred because of its close links with optimal transport. On the one hand, the Schrödinger problem, that comes with a temperature parameter in its classical formulation, converges towards optimal transport when the temperature goes to zero [21, 18, 17]. On the other hand, it is often much easier to compute the solutions of the Schrödinger problem than the ones of the optimal transport problem [8, 23], thanks to the so-called Sinkhorn algorithm [29]. This algorithm converges exponentially fast (*i.e.* at a linear rate, following the usual terminology in the field), at least when it is applied to a reference matrix whose entries are all below bounded by a positive number.

It is known in the theory of matrix scaling that when the reference matrix has nonnegative but possibly cancelling entries, the data in the Schrödinger problem may be chosen in such a way that the latter admits no solution. This is the so-called *non-scalable* case. Also, when the data are located at the boundary of those for which there is a solution, the so-called *approximately scalable* case, the Schrödinger problem has a solution but the convergence of the Sinkhorn algorithm is not linear anymore.

In this paper, we want to study the Sinkhorn algorithm in the degenerate case where the Schrödinger problem has no solution. Our main finding is that for such problem, the Sinkhorn algorithm leads to exactly two limit points, each of them being the solution of a Schrödinger problem with modified data, that we characterize themselves as solutions of auxiliary optimization problems. Also, we show that these limit points are related to a problem where the marginal constraints of the original problem are replaced by marginal penalizations. Moreover, the Schrödinger problem related to the modified data is seen to belong to the approximately scalable case in general. We therefore provide a new outlook on the question of the support of the solution in this case, allowing to design an approximate method for improving the Sinkhorn algorithm's convergence both in the approximately scalable and non-scalable cases.

For simplicity and because it fits with the context of our numerical explorations and needs, we decided to work in finite spaces, even though some of the results might be generalizable.

The Schrödinger problem in finite spaces

Let $\mathcal{D} = \{x_1, \dots, x_N\}$ and $\mathcal{F} = \{y_1, \dots, y_M\}$ be two nonempty finite spaces and $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$ be a nonnegative measure on $\mathcal{D} \times \mathcal{F}$. Of course, we can identify R with a matrix $R = (R_{ij}) \in \mathbb{R}_+^{N\times M}$ by setting $R_{ij} := R(\{(x_i, y_j)\})$. Assuming that R models the coupling between the initial and final positions of the particles of a large system, we interpret R_{ij} as the sum of the masses of all the particles being in x_i at the initial time, and in y_j at the final time.

Let us choose $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$. Once again, we see $\mu = (\mu_i)$ and $\nu = (\nu_j)$ as vectors of $\mathbb{R}_+^N$ and $\mathbb{R}_+^M$ respectively.

We call $\Pi(\mu, \nu)$ the subset of $\mathcal{M}_+(\mathcal{D} \times \mathcal{F})$ consisting of all those matrices $\bar{R}$ whose row and column sums give μ and ν respectively, that is, such that

$$\forall i = 1, \dots, N, \quad \sum_j \bar{R}_{ij} = \mu_i, \qquad \text{and} \qquad \forall j = 1, \dots M, \quad \sum_i \bar{R}_{ij} = \nu_j.$$

In our interpretation, it means that for the system described by $\bar{R}$, the sum of the masses of all the particles being in $x_i \in \mathcal{D}$ at the initial time is μ_i, and the sum of the masses of all the particles being in $y_j \in \mathcal{F}$ at the final time is ν_j. In particular, for $\Pi(\mu, \nu)$ to be nonempty, μ and ν need to share their total mass.

Remark 1. Let us point out to the readers acquainted with the notations used in the Optimal Transport literature that calling $X : \mathcal{D} \times \mathcal{F} \to \mathcal{D}$ and $Y : \mathcal{D} \times \mathcal{F} \to \mathcal{F}$ the canonical projections and denoting by $_\#$ the push forward operation on measures, the measure $\bar{R}$ belongs to $\Pi(\mu, \nu)$ provided $X_\#\bar{R} = \mu$ and $Y_\#\bar{R} = \nu$. Actually, we will not use these notations, and prefer to define $\mu^{\bar{R}} := X_\#\bar{R}$ and $\nu^{\bar{R}} := Y_\#\bar{R}$, see formula 3.

We call the Schrödinger problem w.r.t. R between μ and ν the convex optimization problem consisting in minimizing among $\Pi(\mu, \nu)$ the relative entropy w.r.t R:

$$\mathrm{Sch}(R; \mu, \nu) := \min \Big\{ H(\bar{R}|R) \;\Big|\; \bar{R} \in \Pi(\mu, \nu) \Big\},$$

where for all $\bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, the relative entropy of P w.r.t. R is defined by

$$H(\bar{R}|R) := \sum_{ij} \Big\{ \bar{R}_{ij} \log \frac{\bar{R}_{ij}}{R_{ij}} + R_{ij} - \bar{R}_{ij} \Big\},$$

taking the conventions $a \log \frac{a}{0} = +\infty$ if $a > 0$, and $0 \log 0 = 0 \log \frac{0}{0} = 0$. Notice that if $H(\bar{R}|R) < +\infty$, then for all i, j such that $R_{ij} = 0$, we also have $\bar{R}_{ij} = 0$, *i.e.*, $\bar{R} \ll R$ in the sense of measures.

Remark 2. Of course, as before, for a solution R^* to exist, μ and ν need to have the same total mass, and R^* will then have the same total mass as μ and ν.

By strict convexity of the relative entropy as a function of $\bar{R}$, when there is a solution, the latter is unique. Also, the relative entropy being lower semicontinuous w.r.t. $\bar{R}$ and $\Pi(\mu, \nu)$ being compact, the existence of a solution R^* for $\mathrm{Sch}(R; \mu, \nu)$ is equivalent to the existence of a $\bar{R} \in \Pi(\mu, \nu)$ satisfying $H(\bar{R}|R) < +\infty$. In what follows, such an $\bar{R}$ is called a competitor for $\mathrm{Sch}(R; \mu, \nu)$.

Heuristically, we seek for the measure R^* that is the closest possible to R in the entropic sense while imposing its first and second marginals.

In virtue of the Sanov theorem [25], this problem has an interpretation in terms of large deviations. It is also known to be connected to optimal transport problems, see [18, 17, 11, 5]: if for all i, j, c_{ij} models the cost to transport a unit of mass from x_i to y_j, and $R_{ij} \propto \exp(-c_{ij}/\varepsilon)$ for some small $\varepsilon > 0$, then the solution of $\mathrm{Sch}(R; \mu, \nu)$ is a good approximation of a solution of the optimal transport problem between μ and ν, of cost (c_{ij}).

The Sinkhorn algorithm

When the solution of $\mathrm{Sch}(R; \mu, \nu)$ exists, it is well known for a very long time that this solution turns out to be the limit of the sequences $(P^n)_{n \in \mathbb{N}^*}$ and $(Q^n)_{n \in \mathbb{N}^*}$ appearing in the following so-called the Sinkhorn algorithm, also called IPFP for *iterative proportional fitting procedure* [29, 30, 14, 22]:

$$\begin{cases} Q^0 := R, \\ \forall n \geq 0, \quad P^{n+1} := \arg\min \Big\{ H(P|Q^n), \quad X_\#P = \mu \Big\}, \\ \forall n \geq 0, \quad Q^{n+1} := \arg\min \Big\{ H(Q|P^{n+1}), \quad Y_\#Q = \nu \Big\}. \end{cases} \tag{1}$$

This formulation is implicit as it involves minimization problems. In fact, easy results concerning these problems, detailed in Corollary 6 below, give access to an explicit and easily computable version, which takes the following form, when expressed in terms of the so called *dual variables*

or *potentials* $(a^n)_{n\in\mathbb{N}^*} \in (\mathbb{R}_+^N)^{\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}} \in (\mathbb{R}_+^M)^{\mathbb{N}}$:

$$\begin{cases} \forall j, \quad b_j^0 := 1, \\ \forall n \geq 0, \quad \forall i,j, \quad a_i^{n+1} := \dfrac{\mu_i}{\sum\limits_{j'} b_{j'}^n R_{ij'}}, \qquad P_{ij}^{n+1} := a_i^{n+1} b_j^n R_{ij}, \\ \forall n \geq 0, \quad \forall i,j, \quad b_j^{n+1} := \dfrac{\nu_j}{\sum\limits_{i'} a_{i'}^{n+1} R_{i'j}}, \quad Q_{ij}^{n+1} := a_i^{n+1} b_j^{n+1} R_{ij}. \end{cases} \tag{2}$$

A reason for the popularity of this algorithm is that in a lot of contexts, the sequences of potentials (a^n) and (b^n), and hence the sequence of couplings (P^n) and (Q^n) converge at a linear rate, and the limit of (P^n) and (Q^n) coincide with the unique solution of $\mathrm{Sch}(R;\mu,\nu)$. For this reason, the Sinkhorn algorithm is nowadays the most efficient way to compute approximate solutions of optimal transport problems [8, 2, 23].

Observe that *a priori*, the existence of a solution for the Schrödinger problem is not necessary to give a meaning to the Sinkhorn algorithm. Actually, we will see that there are lots of situations where the Schrödinger problem has no solution, and yet the Sinkhorn algorithm is perfectly well defined. These are the cases that we want to study in this text.

A degenerate case

As we just said, our aim is to study the Sinkhorn algorithm in the cases where the existence of a solution of the Schrödinger problem is either false, or at least nontrivial. This includes the case where μ and ν do not have the same total mass, see Remark 2. However, this is not the main new situation that we want to encompass, since the Sinkhorn algorithm behaves trivially under normalization. More interestingly, we will give a detailed study of the case where some entries of R cancel, or in optimal transport terms, when the cost function takes the value $+\infty$.

In that situation, it can be hard to exhibit a competitor, since the natural candidate that is usually chosen, namely, the product measure of μ and ν, is not absolutely continuous w.r.t. R in general. In fact, there are cases where it is easy to see that no competitor exists. We give in Appendix A an explicit and simple example of such a case. To illustrate our findings, we also describe the behaviour of the Sinkhorn algorithm applied to this example.

Note that beyond the theoretical interest, there are practical motivations for studying cases where the problem has no solution. Indeed, the Schrödinger problem can be used as follows. Suppose that μ and ν are some observed densities of a random phenomenon at two different timepoints, obtained for instance by building the empirical distributions associated to some collected data. Suppose also that we have at our disposal a good model for this phenomenon, that is, a reference stochastic process chosen based on our knowledge of the system prior to the observations of μ and ν. Let us call R the coupling of this process between the two studied timepoints. If we believe enough in our model and in our data, but still the marginals of R are not μ and ν, then it is reasonable to try to improve our model by looking for the coupling that is the closest to R (for instance in the entropic sense), but which is compatible with the data: this means solving $\mathrm{Sch}(R;\mu,\nu)$.

Now imagine that there is no lower-bound for the coupling R, which can be perfectly justified (think for instance of a nondecreasing process, like the size of some randomly growing phenomenon). Then, small measurement errors due to imprecision of the devices or even to too restricted samplings may result in the non-existence of any coupling with marginals μ and ν being absolutely continuous with respect to R: the Schrödinger problem would thus have no solution. In that case, we would like to be able to find a coupling which explains the best

the data while being entropically close to R. Some methods are available for doing so, like for instance algorithms solving the so-called unbalanced problem [6], but at the cost of introducing a new parameter quantifying the balance between the proximity to the data and to the reference coupling, whose value will often be arbitrarily chosen. We show in this article that interestingly, the Sinkhorn algorithm allows to overcome this choice in the specific situation where the data are more trustworthy than the model.

In particular, we were motivated by an application of the Sinkhorn algorithm related to systems biology, and more specifically to the treatment of single-cell data. The quick progresses of acquisition methods for such data raises the hope of a better understanding of the cell-differentiation process, which would in turn pave the way for major medical breakthroughs. In the seminal papers [26, 16], Schiebinger and his coauthors suggest to analyse the collected data through an approach based on optimal transport and more specifically on the Schrödinger problem.

In this field, the unknown is the law of the evolution of the quantity of mRNA molecules in the cells through time: this evolution cannot be followed, as our techniques of measurement destroy the cells. Hence, to study it between two timepoints, the approach consists in:

(i) choosing a reference theoretical model R, where for all i, j, R_{ij} is the expected quantity of cells whose mRNA levels are given by the vector x_i at the initial time, and by y_j at the final time;

(ii) measuring the mRNA levels of samples of cells at the initial and final times to get approximate distributions μ and ν of these levels among the population of cells under study;

(iii) solving the Schrödinger problem $\mathrm{Sch}(R; \mu, \nu)$ to get a law R^* that is close to our model R, but which explains the data.

In the case of Schiebinger, R is the coupling produced by a Brownian motion between two time points, and therefore admits a below bound. In a separated work [31], the second author argues that a more realistic model would be obtained by replacing the Brownian motion by a *piecewise deterministic Markov process* as described in [13]. For such models, dynamical constraints involving mRNAs half-life times lead to a degenerate R and the corresponding Schrödinger problem could thus have no solution, not because of a lack in the model, but because of inaccuracies in the measurements. Our results show that the Sinkhorn algorithm can still be used in this situation, without any pre-treatment of the data. We refer once again to Appendix A for a further discussion on this topic.

Contributions

In this article, we work with a potentially degenerate R, and our main contributions are the following.

- We show that the two sequences $(P^n)_{n\in\mathbb{N}^*}$ and $(Q^n)_{n\in\mathbb{N}^*}$ defined in (1) converge towards two possibly different matrices P^* and Q^*, each of them being the solution of a Schrödinger problem with modified marginals. More precisely, the matrice P^* is the solution of the problem $\mathrm{Sch}(R; \mu, \nu^*)$, where ν^* minimizes the relative entropy w.r.t. ν within the set of marginals $\bar{\nu}$ for which the Schrödinger problem $\mathrm{Sch}(R; \mu, \bar{\nu})$ admits a solution, and a similar statement holds for Q^*. This result, stated at Theorem 11, is the main result of Section 3.

- We show in Section 4 that the Sinkhorn algorithm enables to compute the solution of a modified Schrödinger problem where the marginal *constraints* are replaced by marginal *penalizations*: as shown at Theorem 17, the limit of the solution of the problem

$$\min\Big\{H(\bar{R}|R) + \lambda\big(H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu)\big)\ \Big|\ \bar{R} \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})\Big\},$$

(where once again, $\mu^{\bar{R}}$ and $\nu^{\bar{R}}$ are the first and second marginal of $\bar{R}$, see Remark 1) converges towards the componentwise geometric mean of the two limits P^* and Q^* of the Sinkhorn algorithm as $\lambda \to +\infty$.

- In Section 5, we recall a well known necessary and sufficient condition on R, μ and ν for $\mathrm{Sch}(R;\mu,\nu)$ to admit a solution. Using this condition, we develop at Proposition 28 a procedure to find the (common) support $\mathcal{S}$ of P^* and Q^* without computing them. As explained in Subsection 5.2, our motivation is that the convergence rate of the Sinkhorn algorithm is linear if and only if $\mathcal{S}$ coincides with the support of R. When it is not the case, as often when $\mathrm{Sch}(R;\mu,\nu)$ has no solution, we can therefore improve the speed of convergence by first computing $\mathcal{S}$, and then by applying the Sinkhorn algorithm to $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}}R;\mu,\nu)$ instead of $\mathrm{Sch}(R;\mu,\nu)$, which does not change the limits P^* and Q^*.

- Section 6 is an application of the developments made at Section 5. We implement an approximate but fast algorithm, usable in practice, allowing to recover an estimate of the support $\mathcal{S}$. We then compare the Sinkhorn algorithm and the technique coming from [6] with our method consisting in first computing $\mathcal{S}$ with our approximate algorithm and then applying the Sinkhorn algorithm to $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}}R;\mu,\nu)$. We also detail the regimes in which our method is a significant improvement of the other techniques.

Some of the results of this paper can be generalized by replacing $\mathcal{D}$ and $\mathcal{F}$ by general Polish spaces without much effort. This is the reason why we will often write $H(P|R) < +\infty$ instead of $P \ll R$: these are equivalent in the finite case, but not in the continuous one. In the latter case, we often need the stronger entropic assumption. Even if we decided to stick to the finite case in order to stress the key arguments that make everything work in practice, we believe that the continuous case is also interesting, and we wish to study it in a further work.

Before coming up with our contributions, we recall a few facts about the relative entropy functional at Section 2.

2 Notations, properties of the entropy and terminology

In this preliminary section, we introduce some notations, provide well known elementary results concerning the entropy, and recall the terminology usually used in the theory of matrix scaling.

2.1 Notations

Let us first give a few notations that will be used systematically in this work. Most of them were already given in the introduction.

- Whenever I is a finite set of labels and $\mathcal{E} = \{u_k,\ k \in I\}$ is a finite set indexed by I, we denote by $\mathcal{M}_+(\mathcal{E})$ the set of nonnegative measures on $\mathcal{E}$. This set is identified with $\mathbb{R}_+^I$ through the the correspondence

$$\mathsf{r} \in \mathcal{M}_+(\mathcal{E}) \leftrightsquigarrow (\mathsf{r}_k := \mathsf{r}(\{u_k\}))_{k\in I} \in \mathbb{R}^I.$$

  For all $\mathsf{r} \in \mathcal{M}_+(\mathcal{E})$, we denote by $\mathsf{M}(\mathsf{r}) := \sum_k \mathsf{r}_k$ its total mass. If $\mathsf{M}(\mathsf{r}) = 1$, we say that r is a probability measure on $\mathcal{E}$, and we write $\mathsf{r} \in \mathcal{P}(\mathcal{E})$. The topology considered on $\mathcal{M}_+(\mathcal{E})$ is nothing but the one of $\mathbb{R}^I$.

- In the same way, we identify the set $\mathcal{F}(\mathcal{E};\mathbb{R})$ of real functions Z on $\mathcal{E}$ with $\mathbb{R}^I$ through the correspondence

$$Z \in \mathcal{F}(\mathcal{E};\mathbb{R}) \leftrightsquigarrow (Z_k := Z(u_k))_{k\in I} \in \mathbb{R}_+^I.$$

Depending on the context, we will either call such functions Z *test functions*, or *random variables*, thinking of $\mathcal{E}$ as a measurable set. The random variables that we will consider will actually often be slightly more general, and be allowed to take the value $-\infty$, in which case we will tell it explicitly.

- Through our identifications, the duality between $\mathcal{M}_+(\mathcal{E})$ and $\mathcal{F}(\mathcal{E};\mathbb{R})$ is nothing but the usual scalar product on $\mathbb{R}^I$, and denoted for all $Z \in \mathcal{F}(\mathcal{E};\mathbb{R})$ and $\mathsf{r} \in \mathcal{M}_+(\mathcal{E})$ by

$$\langle Z, \mathsf{r} \rangle := \sum_k Z_k \mathsf{r}_k.$$

When Z possibly takes the value $-\infty$, we always choose by convention $-\infty \times 0 = 0$.

- In the context of the introduction, when $\mathcal{D} = \{x_1, \dots, x_N\}$ and $\mathcal{F} = \{y_1, \dots, y_M\}$ are two nonempty finite spaces and $\mathcal{E} = \mathcal{D} \times \mathcal{F}$, then the corresponding I is the product space $\{1, \dots, N\} \times \{1, \dots, M\}$, and $\bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, is seen as a matrix. We define its marginals $\mu^{\bar{R}} \in \mathcal{M}_+(\mathcal{D})$ and $\nu^{\bar{R}} \in \mathcal{M}_+(\mathcal{F})$ by the formulas

$$\forall i = 1, \dots, N, \quad \mu_i^{\bar{R}} := \sum_j \bar{R}_{ij}, \quad \text{and} \quad \forall j = 1, \dots M, \quad \nu_j^{\bar{R}} := \sum_i \bar{R}_{ij}. \tag{3}$$

Of course, $\mu^{\bar{R}}$ and $\nu^{\bar{R}}$ have the same total mass as $\bar{R}$, that is:

$$\mathsf{M}(\bar{R}) = \mathsf{M}(\mu^{\bar{R}}) = \mathsf{M}(\nu^{\bar{R}}). \tag{4}$$

In particular, if R is a probability measure, its marginals are probability measures as well.

- As before, if $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$, we call $\Pi(\mu, \nu)$ the set of those $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$ such that $\mu^R = \mu$ and $\nu^R = \nu$.

- For the sake of simplicity, we do not use different notations for the same functions applied in different context. For instance, notations for the total mass M or the relative entropy H (see Definition 3 below) might be applied to different sets $\mathcal{E}$ namely $\mathcal{D}$, $\mathcal{F}$ and $\mathcal{D} \times \mathcal{F}$.

2.2 First properties of the relative entropy

This subsection only contains easy and very well known results concerning the relative entropy that will be useful in the sequel. We stick to the finite case as this is the one studied in this paper, and we provide some proofs for the readers who are not acquainted with this notion of entropy, but all the properties given here are known in a much wider context, see for instance [19].

As already said in the introduction, the relative entropy is defined as follows.

Definition 3. Let $\mathcal{E} = \{u_k,\ k \in I\}$ be a finite set and $\mathsf{r} = (\mathsf{r}_k) \in \mathcal{M}_+(\mathcal{E})$. For all $\bar{\mathsf{r}} = (\bar{\mathsf{r}}_k) \in \mathcal{M}_+(\mathcal{E})$, the relative entropy of $\bar{\mathsf{r}}$ w.r.t r is the value in $[0, +\infty]$ given by

$$H(\bar{\mathsf{r}}|\mathsf{r}) := \sum_k \left\{ \bar{\mathsf{r}}_k \log \frac{\bar{\mathsf{r}}_k}{\mathsf{r}_k} + \mathsf{r}_k - \bar{\mathsf{r}}_k \right\} = \sum_k \bar{\mathsf{r}}_k \log \frac{\bar{\mathsf{r}}_k}{\mathsf{r}_k} + \mathsf{M}(\mathsf{r}) - \mathsf{M}(\bar{\mathsf{r}}),$$

with convention $a \log \frac{a}{0} = +\infty$ for all $a > 0$, and $0 \log 0 = 0 \log \frac{0}{0} = 0$.

First, this definition provides a convex function with good continuity properties. We state them in the following proposition, for which we omit the straightforward proof.

Proposition 4. *Let $\mathcal{E}$ be a finite set and $\mathsf{r} \in \mathcal{M}_+(\mathcal{E})$. The functional*

$$\bar{\mathsf{r}} \in \mathcal{M}_+(\mathcal{E}) \mapsto H(\bar{\mathsf{r}}|\mathsf{r}) \in [0, +\infty]$$

is strictly convex, lower semicontinuous, and continuous on its domain, which is the closed set $\{\bar{\mathsf{r}} \ll \mathsf{r}\} \subset \mathcal{M}_+(\mathcal{E})$.

For a given $\bar{\mathsf{r}} \in \mathcal{M}_+(\mathcal{E})$*, the functional*

$$\mathsf{r} \in \mathcal{M}_+(\mathcal{E}) \mapsto H(\bar{\mathsf{r}}|\mathsf{r}) \in [0, +\infty]$$

is convex and continuous for the canonical topology of $[0, +\infty]$*. Its domain is the open set* $\{\mathsf{r} \gg \bar{\mathsf{r}}\} \subset \mathcal{M}_+(\mathcal{E})$.

The most useful property of the relative entropy is the computation of its Legendre transform. This property can be stated as follows.

Theorem 5. *Let* $\mathcal{E} = \{u_k,\ k \in I\}$ *be a finite set, and* $\mathsf{r} \in \mathcal{M}_+(\mathcal{E})$*. For all test function* Z *possibly taking the value* $-\infty$ *on* $\mathcal{E}$ *and all nonnegative measure* $\bar{\mathsf{r}}$ *on* $\mathcal{E}$*, we have*

$$\langle Z, \bar{\mathsf{r}} \rangle \leq H(\bar{\mathsf{r}}|\mathsf{r}) + \langle e^Z - 1, \mathsf{r} \rangle, \tag{5}$$

with conventions $e^{-\infty} = 0$*,* $-\infty \times 0 = 0$ *and* $+\infty - \infty = +\infty$.

Moreover, equality in $\mathbb{R}$ *holds if and only if* $\bar{\mathsf{r}} \ll \mathsf{r}$ *and for all* $k \in I$*,*

$$Z_k = \log \frac{\bar{\mathsf{r}}_k}{\mathsf{r}_k} \in [-\infty, +\infty) \tag{6}$$

with convention $\log \frac{0}{a} = -\infty$ *for all* $a \geq 0$.

Proof. Let $\mathsf{r}, \bar{\mathsf{r}}$ and Z be as in the statement of the theorem. If $H(\bar{\mathsf{r}}|\mathsf{r}) = +\infty$, there is nothing to prove, and we assume $\bar{\mathsf{r}} \ll \mathsf{r}$.

By direct real computations, with the same conventions as in the statement of the theorem, we find that for all $k \in I$:

$$Z_k \bar{\mathsf{r}}_k \leq \bar{\mathsf{r}}_k \log \frac{\bar{\mathsf{r}}_k}{\mathsf{r}_k} + \mathsf{r}_k - \bar{\mathsf{r}}_k + \big(e^{Z_k} - 1\big)\mathsf{r}_k,$$

with equality if and only if $\mathsf{r}_k = \bar{\mathsf{r}}_k = 0$ or $\mathsf{r}_k > 0$ and

$$Z_k = \log \frac{\bar{\mathsf{r}}_k}{\mathsf{r}_k} \in [-\infty, +\infty).$$

We find (5) and (6) by summing this inequality over k. □

This theorem will be useful as such, but also implies the following corollary which gives a full understanding of one step in the Sinkhorn algorithm (1).

Corollary 6. *Let* $\mathcal{D}$ *and* $\mathcal{F}$ *be two finite sets, and* $\bar{R}, R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$*. With the notations of* (3)*, we have*

$$H(\mu^{\bar{R}}|\mu^R) \leq H(\bar{R}|R) \qquad \text{and} \qquad H(\nu^{\bar{R}}|\nu^R) \leq H(\bar{R}|R). \tag{7}$$

In the case where $H(\bar{R}|R)$ *is finite, equality holds if and only if for all* i, j*, respectively:*

$$\bar{R}_{ij} = \frac{\mu_i^{\bar{R}}}{\mu_i^R} R_{ij} \qquad \text{and} \qquad \bar{R}_{ij} = \frac{\nu_j^{\bar{R}}}{\nu_j^R} R_{ij},$$

with convention $\frac{0}{0} = 0$.

In particular, given $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$ *and* $\mu \in \mathcal{M}_+(\mathcal{D})$*, the problem*

$$\min \Big\{ H(P|R) \,\Big|\, \mu^P = \mu \Big\} \tag{8}$$

admits a solution if and only if $H(\mu|\mu^R) < +\infty$*, and in this case, this solution* P *is unique and satisfies for all* i, j

$$P_{ij} = \frac{\mu_i}{\mu_i^R} R_{ij} \tag{9}$$

with convention $\frac{0}{0} = 0$*. Moreover,* $H(P|R) = H(\mu|\mu^R)$*.*

Similarly, given $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$ *and* $\nu \in \mathcal{M}_+(\mathcal{F})$*, the problem*

$$\min \left\{ H(Q|R) \,\middle|\, \nu^Q = \nu \right\}$$

admits a solution if and only if $H(\nu|\nu^R) < +\infty$*, and in this case, this solution* Q *is unique and satisfies for all* i, j

$$Q_{ij} = \frac{\nu_j}{\nu_j^R} R_{ij}$$

with convention $\frac{0}{0} = 0$*. Moreover,* $H(Q|R) = H(\nu|\nu^R)$*.*

Proof. The first inequality in (7) is a direct application of (5) with $\mathsf{r} = R$, $\bar{\mathsf{r}} = \bar{R}$ and for all i, j,

$$Z_{ij} = \log \frac{\mu_i^{\bar{R}}}{\mu_i^R}.$$

The second inequality is proved in the same way, and the equality case is a consequence of (6).

For the second part of the statement, let us observe that for all P satisfying the constraint in (8), because of (7), $H(P|R) \geq H(\mu|\mu^R)$, which – by the equality case – is attained if and only if (9) holds. The problem involving the second marginal is treated in the same way. □

2.3 The Schrödinger problem: assumptions and terminology

Let $\mathcal{D} = \{x_1, \dots, x_N\}$ and $\mathcal{F} = \{y_1, \dots, y_M\}$ be two nonempty finite sets, and let us choose a reference measure $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$. Given $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$, the Schrödinger problem, already defined in the introduction, rewrites with the notations of Subsection 2.1:

$$\mathrm{Sch}(R; \mu, \nu) := \min \left\{ H(\bar{R}|R) \,\middle|\, \bar{R} \in \mathcal{M}_+(\mu, \nu) \text{ such that } \mu^{\bar{R}} = \mu \text{ and } \nu^{\bar{R}} = \nu \right\}. \tag{10}$$

Remark 7. Here, we define $\mathrm{Sch}(R; \mu, \nu)$ as the optimal value of our problem. However, with an abusive terminology, we will refer to the minimizer of the r.h.s. of (10) as "the solution of $\mathrm{Sch}(R; \mu, \nu)$". More generally, we will call "the problem $\mathrm{Sch}(R; \mu, \nu)$" the optimization problem consisting in computing the value $\mathrm{Sch}(R; \mu, \nu)$.

As we will see in Theorem 11, the Sinkhorn algorithm (1) associated with the problem $\mathrm{Sch}(R; \mu, \nu)$ is well defined if and only if the following assumption holds.

Assumption 8. *Let* $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$*,* $\mu \in \mathcal{M}_+(\mathcal{D})$ *and* $\nu \in \mathcal{M}_+(\mathcal{F})$*, and let us call*

$$\mathcal{E} := \left\{ (x_i, y_j) \in \mathcal{D} \times \mathcal{F} \text{ such that } R_{ij} > 0,\ \mu_i > 0 \text{ and } \nu_j > 0 \right\}. \tag{11}$$

We say that the triple $(R; \mu, \nu)$ *satisfies Assumption 8 provided* $R^0 := \mathbb{1}_{\mathcal{E}} \cdot R$ *is such that:*

$$H(\mu|\mu^{R^0}) < +\infty \qquad \text{and} \qquad H(\nu|\nu^{R^0}) < +\infty. \tag{12}$$

This assumption is easily seen to be necessary for $\mathrm{Sch}(R; \mu, \nu)$ to admit a solution. Under Assumption 8 either $\mathsf{M}(\mu) = \mathsf{M}(\nu) = 0$, or none of them is 0. In the second case, up to replacing $\mathcal{D}$ by $\mathcal{D}'$, the support of μ, $\mathcal{F}$ by $\mathcal{F}'$, the support of ν, and R by its restriction (or equivalently of the one of R^0) on $\mathcal{D}' \times \mathcal{F}'$, we end up with the following assumption, that will often be used in this paper.

Assumption 9. *Let $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$. We say that the triple $(R; \mu, \nu)$ satisfies Assumption 9 provided the support of μ and μ^R is $\mathcal{D}$ and the support of ν and ν^R is $\mathcal{F}$.*

The Schrödinger problem (10) consists in minimizing a convex function under linear constraints. Therefore, the functional $(\mu, \nu) \in \mathcal{M}_+(\mathcal{D}) \times \mathcal{M}_+(\mathcal{F}) \mapsto \mathrm{Sch}(R; \mu, \nu) \in [0, +\infty]$ is convex.

In the case where Assumption 9 holds, following the usual terminology of the matrix scaling theory (except for the last item which is more exotic), see [14], we say that:

- The problem is *scalable* if (μ, ν) is in the relative interior of the domain of $\mathrm{Sch}(R; \cdot)$. In this case, $\mathsf{M}(\mu) = \mathsf{M}(\nu)$, the Schrödinger problem admits a unique solution R^*, $R^* \sim R$ in the sense of measures, and the Sinkhorn algorithm converges towards R^*, at a linear rate. In Lemma 24, we recall an explicit necessary and sufficient condition on R, μ, ν for $\mathrm{Sch}(R; \mu, \nu)$ to be scalable.

- The problem is *approximately scalable* if (μ, ν) is at the relative boundary of the domain of $\mathrm{Sch}(R; \cdot)$. In this case, $\mathsf{M}(\mu) = \mathsf{M}(\nu)$, the Schrödinger problem admits a unique solution R^*, and the Sinkhorn algorithm converges towards R^*. However, in this case, the support of R^* is strictly included in the support of R (else, we easily see that we are in the scalable case), and the rate cannot be linear anymore: as proved in [1], a linear rate of convergence for the Sinkhorn algorithm is not compatible with the appearance of new zero entries at the limit. We recall at Theorem 23 a necessary and sufficient condition on R, μ and ν for $\mathrm{Sch}(R; \mu, \nu)$ to be at least approximately scalable, that is, either approximately scalable or scalable.

- The problem is *non-scalable* if $\mathsf{M}(\mu) = \mathsf{M}(\nu)$, but the Schrödinger problem $\mathrm{Sch}(R; \mu, \nu)$ does not admit a solution. This is the case when the condition of Theorem 23 does not hold. This case is the main case of interest in this work.

- The problem is *unbalanced* if $\mathsf{M}(\mu) \neq \mathsf{M}(\nu)$. Calling $\mu' := \mu/\mu(\mathcal{D})$ and $\nu' := \nu/\nu(\mathcal{F})$ their normalized versions, we will say that $\mathrm{Sch}(R; \mu, \nu)$ is respectively unbalanced scalable, unbalanced approximately scalable and unbalanced non-scalable whenever $\mathrm{Sch}(R; \mu', \nu')$ is scalable, approximately scalable or non-scalable.

Yet, with an abuse of terminology, we will often refer to the non-scalable case for results that are true in *any* situation, including the balanced and unbalanced non-scalable ones, which are often the most difficult.

3 The Sinkhorn algorithm in the non-scalable case

In this section, we consider $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ that we identify respectively with a matrix and two vectors, as before.

The goal of this section is to show that under obvious necessary assumptions, then the algorithm given in (1) is well defined, and that the sequences $(P^n)_{n \in \mathbb{N}^*}$ and $(Q^n)_{n \in \mathbb{N}^*}$ that it provides converge separately towards matrices P^* and Q^* that we define now. It will be obvious from their definition that these matrices coincide if and only if the problem $\mathrm{Sch}(R; \mu, \nu)$ defined in (10) admits a solution, that is, if it is at least approximately scalable. Hence our proof recovers the classical fact that the Sinkhorn algorithm converges towards the solution of the Schrödinger problem as soon as the latter exists.

The first step to define P^* and Q^* is to define a pair of new marginals $\mu^* \in \mathcal{M}_+(\mathcal{D})$ and $\nu^* \in \mathcal{M}_+(\mathcal{F})$ as solutions of the following optimization problem:

$$\begin{aligned} \mu^* &:= \arg\min \Big\{ H(\bar{\mu}|\mu) \,\Big|\, \bar{\mu} = \mu^Q \text{ for some } Q \text{ with } H(Q|R) < +\infty \text{ and } \nu^Q = \nu \Big\}, \\ \nu^* &:= \arg\min \Big\{ H(\bar{\nu}|\nu) \,\Big|\, \bar{\nu} = \nu^P \text{ for some } P \text{ with } H(P|R) < +\infty \text{ and } \mu^P = \mu \Big\}. \end{aligned} \tag{13}$$

The question of existence of μ^* and ν^* is treated in Theorem 11 below. Of course, if the problem $\mathrm{Sch}(R;\mu,\nu)$ admits a competitor, then $\mu^* = \mu$ and $\nu^* = \nu$.

Remark 10. In the unbalanced case, notice that the total mass of ν^* is the one of μ, and the total mass of μ^* is the one of ν, that is, $\mathsf{M}(\nu^*) = \mathsf{M}(\mu)$ and $\mathsf{M}(\mu^*) = \mathsf{M}(\nu)$.

Then P^* and Q^* are simply defined as the solutions of the Schrödinger problems $\mathrm{Sch}(R;\mu,\nu^*)$ and $\mathrm{Sch}(R;\mu^*,\nu)$ respectively, that is:

$$P^* := \arg\min \Big\{ H(P|R) \,\Big|\, P \in \Pi(\mu,\nu^*) \Big\} \quad \text{and} \quad Q^* := \arg\min \Big\{ H(Q|R) \,\Big|\, Q \in \Pi(\mu^*,\nu) \Big\}. \tag{14}$$

Of course, if the problem $\mathrm{Sch}(R;\mu,\nu)$ admits a competitor, and hence a solution, then both P^* and Q^* coincide with this solution.

Our convergence theorem can be stated as follows.

Theorem 11. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfy Assumption 8. The sequences $(P^n)_{n\in\mathbb{N}^*}$ and $(Q^n)_{n\in\mathbb{N}^*}$ from (1), the marginals μ^* and ν^* from (13) and the matrices P^* and Q^* from (14) are well defined, and*

$$P^n \underset{n\to+\infty}{\longrightarrow} P^* \qquad \text{and} \qquad Q^n \underset{n\to+\infty}{\longrightarrow} Q^*.$$

Remark 12.

- Assumption 8 is necessary: it is straightforward to check that if Q^1 from (1) is well defined, then $H(Q^1|R^0) < +\infty$. In particular, projecting on the second marginal, we conclude that $H(\nu|\nu^{R^0}) < +\infty$. Arguing in the same way with P^2 in place of Q^1 and the second marginal in place of the first one, we see that if P^2 is well defined, then $H(\mu|\mu^{R^0}) < +\infty$. In particular, there is nothing to check before starting the algorithm: if the algorithm is able to compute P^2, then it means that our assumption is satisfied and that the convergence holds.

- Note that the topology for the convergence stated in the theorem does not matter since we are working in finite dimensional spaces. However, we believe that the result is still true replacing $\mathcal{D}$ and $\mathcal{F}$ by general Polish spaces. In this case, the convergence needs to be understood in the sense of the narrow topology, a topology for which the sequences (P^n) and (Q^n) can be proved to be compact due to the properties of their marginals.

- Remarkably, we will be able to prove this theorem without deriving the optimality conditions for μ^* and ν^*. However, these optimality conditions will be needed in the next section, and hence written at Proposition 19.

- As developed in [7], there exists a strong analogy between the relative entropy and square Euclidean distances, and this in spite of the lack of symmetry of the first. In particular, following this analogy, the Sinkhorn algorithm (1) consists in iteratively othogonaly projecting on the convex sets of measures absolutely continuous w.r.t. R satisfying the first and second marginal constraint respectively.

  With this picture in mind, we can give in Figure 1 a visual representation of the scalable and non-scalable case. In the scalable case, the two convex sets intersect, and the sequences $(P^n)_{n\in\mathbb{N}^*}$ and $(Q^n)_{n\in\mathbb{N}^*}$ converge towards the point of the intersection that is

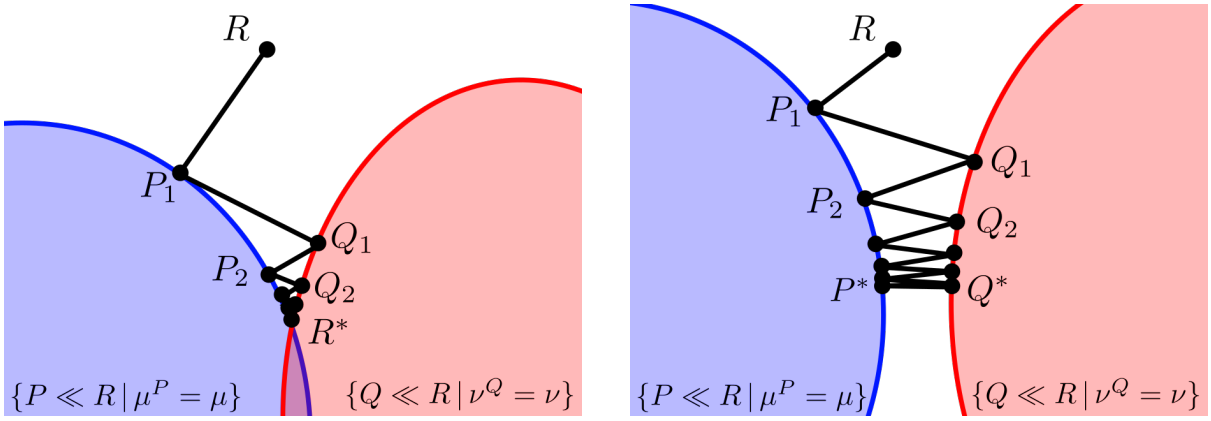


Figure 1 : Sketchy representation of the Sinkhorn algorithm in the scalable case (to the left) and nonscalable case (to the right).

the closest to R. In the non-scalable case, the two convex sets do not intersect. However, the sequences $(P^n)_{n\in\mathbb{N}^*}$ and $(Q^n)_{n\in\mathbb{N}^*}$ still converge respectively to P^* and Q^*, the two extreme points of the shortest line segment connecting both sets. Theorem 11 indeed justifies this type of behaviour for the Sinkhorn algorithm.

One should still keep in mind that this analogy and our drawings are only sketchy. In reality, the projections are not orthogonal, and the convex sets have polygonal borders.

Proof. Step 1: All the objects are well defined.

Let us first show that under the assumption of the theorem, the sequences $(P^n)_{n\in\mathbb{N}^*}$ and $(Q^n)_{n\in\mathbb{N}^*}$ are well defined. We start with P^1. As by assumption $\mu \ll \mu^{R^0} \ll \mu^R$, Corollary 6 shows that P^1 is well defined, and that for all i, j,

$$P^1_{ij} = \frac{\mu_i}{\mu^R_i} R_{ij},$$

with convention $\frac{0}{0} = 0$. Clearly, $R^0 \ll P^1$, as the support of the latter is

$$\Big\{(x_i, y_j) \in \mathcal{D} \times \mathcal{F} \text{ s.t. } R_{ij} > 0 \text{ and } \mu_i > 0\Big\} \supset \mathcal{E}.$$

Therefore, $\nu \ll \nu^{R^0} \ll \nu^{P^1}$. So once again, Corollary 6 shows that Q^1 is well defined, and that for all i, j,

$$Q^1_{ij} = \frac{\nu_j}{\nu^{P^1}_j} P^1_{ij},$$

with convention $\frac{0}{0} = 0$. The support of Q^1 is

$$\Big\{(x_i, y_j) \in \mathcal{D} \times \mathcal{F} \text{ s.t. } P^1_{ij} > 0 \text{ and } \nu_j > 0\Big\}$$
$$= \Big\{(x_i, y_j) \in \mathcal{D} \times \mathcal{F} \text{ s.t. } R_{ij} > 0 \text{ and } \mu_i > 0 \text{ and } \nu_j > 0\Big\} = \mathcal{E}.$$

Then, a direct induction argument relying on the following formulas holding for all $n \in \mathbb{N}$ and all i, j:

$$P^{n+1}_{ij} = \frac{\mu_i}{\mu^{Q^n}_i} Q^n_{ij} \qquad \text{and} \qquad Q^{n+1}_{ij} = \frac{\nu_i}{\nu^{P^{n+1}}_j} P^{n+1}_{ij}, \tag{15}$$

with convention $\frac{0}{0} = 0$ show that for all $n \geq 2$, P^n and Q^n are well defined and admit $\mathcal{E}$ as their common support.

Let us now show that μ^* and ν^* are well defined. Their role are symmetric, so we just need to show that μ^* is well defined. First Q^1 satisfies $\nu^{Q^1} = \nu$ and $H(Q^1|R) < +\infty$. Therefore, the problem

$$\inf \Big\{ H(\bar{\mu}|\mu),\ \bar{\mu} = \mu^Q \text{ for some } Q \text{ with } H(Q|R) < +\infty \text{ and } \nu^Q = \nu \Big\}$$

consists in minimizing the continuous (on its domain) and strictly convex function $\bar{\mu} \mapsto H(\bar{\mu}|\mu)$ over the nonempty compact convex set

$$\Big\{ \bar{\mu} = \mu^Q \text{ for some } Q \text{ with } H(Q|R) < +\infty \text{ and } \nu^Q = \nu \Big\}.$$

Hence, it admits a unique solution μ^*.

Finally, let us show that P^* and Q^* are well defined. Once again, their role are symmetric, so we only show the existence of Q^*. We already saw that μ^* is well defined. By definition of the latter, there exists $\bar{Q}$ with $\nu^{\bar{Q}} = \nu$, $\mu^{\bar{Q}} = \mu^*$ and $H(\bar{Q}|R) < +\infty$. So $\mathrm{Sch}(R; \mu^*, \nu)$ consists in minimizing the continuous (on its domain) and strictly convex function $Q \mapsto H(Q|R)$ on the nonempty compact convex set

$$\Pi(\mu^*, \nu) \cap \Big\{ Q \in \mathcal{P}(\mathcal{D} \times \mathcal{F}) \text{ such that } H(Q|R) < +\infty \Big\}.$$

So it admits a unique solution Q^*.

Step 2: A formula for $H(Q|R)$, for all $Q \in \Pi(\mu^*, \nu)$ with $H(Q|R) < +\infty$.

Recalling $Q^0 = R$, we infer from (15) that for all $(x_i, y_j) \in \mathcal{E}$ and $n \in \mathbb{N}^*$,

$$Q^n_{ij} = \frac{\nu_j}{\nu_j^{P^n}} \times \frac{\mu_i}{\mu_i^{Q^{n-1}}} \times \cdots \times \frac{\nu_j}{\nu^{P^1_j}} \times \frac{\mu_i}{\mu_i^{Q^0}} \times R_{ij}. \tag{16}$$

Observe that in the product in the r.h.s., because we assumed that $(x_i, y_j) \in \mathcal{E}$, the common support of all the iterates of the Sinkhorn algorithm, all the factors are positive.

In addition, for all $Q \in \Pi(\mu^*, \nu)$ with finite entropy w.r.t. R, the support of Q is included in $\mathcal{E}$. This is because $H(\mu^*|\mu) < +\infty$, and thereby $\mu^* \ll \mu$. Therefore, we deduce that $Q \ll Q^n$.

So as a consequence of (16), for all i, j in the support of Q,

$$\begin{aligned}
\log \frac{Q_{ij}}{R_{ij}} &= \log \frac{Q_{ij}}{Q^n_{ij}} + \sum_{k=1}^n \Big\{ \log \frac{\nu_j}{\nu_j^{P^k}} + \log \frac{\mu_i}{\mu_i^{Q^{k-1}}} \Big\} \\
&= \log \frac{Q_{ij}}{Q^n_{ij}} + \sum_{k=1}^n \Big\{ \log \frac{\nu_j}{\nu_j^{P^k}} - \log \frac{\mu_i^*}{\mu_i} + \log \frac{\mu_i^*}{\mu_i^{Q^{k-1}}} \Big\}.
\end{aligned}$$

Let us multiply this equality by Q_{ij}, and sum over i, j. We get

$$\sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{R_{ij}} = \sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{Q^n_{ij}} + \sum_{k=1}^n \left\{ \sum_{i,j} Q_{ij} \log \frac{\nu_j}{\nu_j^{P^k}} - \sum_{i,j} Q_{ij} \log \frac{\mu_i^*}{\mu_i} + \sum_{i,j} Q_{ij} \log \frac{\mu_i^*}{\mu_i^{Q^{k-1}}} \right\}.$$

Fix k in $\{1, \dots, n\}$, and consider the term:

$$\sum_{i,j} Q_{ij} \log \frac{\nu_j}{\nu_j^{P^k}} = \sum_j \left(\sum_i Q_{ij} \right) \log \frac{\nu_j}{\nu_j^{P^k}}.$$

As the second marginal of Q is ν, we find

$$\sum_{i,j} Q_{ij} \log \frac{\nu_j}{\nu_j^{P^k}} = \sum_j \nu_j \log \frac{\nu_j}{\nu_j^{P^k}}.$$

As the first marginal of Q is μ^*, we can reason in the same way for the other terms of the same type, and find

$$\sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{R_{ij}} = \sum_{i,j} Q_{ij} \log \frac{Q_{ij}}{Q^n_{ij}} + \sum_{k=1}^{n} \left\{ \sum_j \nu_j \log \frac{\nu_j}{\nu_j^{P^k}} - \sum_i \mu_i^* \log \frac{\mu_i^*}{\mu_i} + \sum_i \mu_i^* \log \frac{\mu_i^*}{\mu_i^{Q^{k-1}}} \right\}.$$

We now use the Definition 3 of the relative entropy to find that this identity means:

$$\begin{aligned} H(Q|R) - \mathsf{M}(R) + \mathsf{M}(Q) = H(Q|Q^n) - \mathsf{M}(Q^n) + \mathsf{M}(Q) \\ + \sum_{k=1}^{n} \Big\{ H(\nu|\nu^{P^k}) - \mathsf{M}(\nu) + \mathsf{M}(\nu^{P^k}) - H(\mu^*|\mu) + \mathsf{M}(\mu^*) - \mathsf{M}(\mu) \\ + H(\mu^*|\mu^{Q^{k-1}}) - \mathsf{M}(\mu^*) + \mathsf{M}(\mu^{Q^{k-1}}) \Big\}, \end{aligned}$$

or, simplifying the masses $\mathsf{M}(Q)$ and $\mathsf{M}(\mu^*)$ appearing several times,

$$\begin{aligned} H(Q|R) - \mathsf{M}(R) = H(Q|Q^n) - \mathsf{M}(Q^n) + \sum_{k=1}^{n} \Big\{ H(\nu|\nu^{P^k}) - \mathsf{M}(\nu) + \mathsf{M}(\nu^{P^k}) - H(\mu^*|\mu) - \mathsf{M}(\mu) \\ + H(\mu^*|\mu^{Q^{k-1}}) + \mathsf{M}(\mu^{Q^{k-1}}) \Big\}. \end{aligned}$$

Let us check how the masses simplify. By (4) as Q and Q^k, $k \in \mathbb{N}^*$ admit ν as their second marginals, they have the same total masses, and it coincides with the one of their first marginals. Namely,

$$\forall k \geq 1, \qquad \mathsf{M}(\nu) = \mathsf{M}(Q^k) = \mathsf{M}(Q) = \mathsf{M}(\mu^{Q^k}) = \mathsf{M}(\mu^*).$$

In the same way,

$$\forall k \geq 1, \qquad \mathsf{M}(\mu) = \mathsf{M}(P^k) = \mathsf{M}(\nu^{P^k}).$$

And finally, as $Q^0 = R$, we also have

$$\mathsf{M}(R) = \mathsf{M}(Q^0) = \mathsf{M}(\mu^{Q^0}).$$

Coming back to our entropy identity, we can simplify more and get

$$H(Q|R) = H(Q|Q^n) + \sum_{k=1}^{n} \Big\{ H(\nu|\nu^{P^k}) - H(\mu^*|\mu) \Big\} + \sum_{k=1}^{n} H(\mu^*|\mu^{Q^{k-1}}). \tag{17}$$

(The only subtlety is that for $k = 1$ and for $k = 1$ only, $\mathsf{M}(\mu^{Q^{k-1}})$ does not simplify with $\mathsf{M}(\nu)$. But then $\mathsf{M}(\mu^{Q^{k-1}}) = \mathsf{M}(\mu^{Q^0})$ simplifies with $\mathsf{M}(R)$, and $\mathsf{M}(\nu)$ simplifies with $\mathsf{M}(Q^n)$.)

Now, we claim that every term in the first sum in the r.h.s. is nonnegative, *i.e.* that $H(\nu|\nu^{P^k}) \geq H(\mu^*|\mu)$. First, we know that $\nu^{Q^k} = \nu$, so that:

$$\begin{aligned} H(\nu|\nu^{P^k}) &= \sum_j \nu_j \log \frac{\nu_j}{\nu_j^{P^k}} + \mathsf{M}(\nu^{P^k}) - \mathsf{M}(\nu) \\ &= \sum_{ij} Q^k_{ij} \log \frac{\nu_j}{\nu_j^{P^k}} + \mathsf{M}(\nu^{P^k}) - \mathsf{M}(\nu) \\ &= \sum_{ij} Q^k_{ij} \log \frac{Q^k_{ij}}{P^k_{ij}} + \mathsf{M}(P^k) - \mathsf{M}(Q^k) = H(Q^k|P^k), \end{aligned}$$

where we used at the third line that from (15), we know that for all j and $1 \leq k \leq n$, $\nu_j/\nu_j^{P^k} = Q^k_{ij}/P^k_{ij}$. Second, by Corollary 6,

$$H(\nu|\nu^{P^k}) = H(Q^k|P^k) \geq H(\mu^{Q^k}|\mu^{P^k}) = H(\mu^{Q^k}|\mu).$$

Finally, Q^k has finite entropy w.r.t. R (use for instance (16) with $n = k$) and admits ν as a first marginal. So by optimality of μ^*, $H(\mu^{Q^k}|\mu) \geq H(\mu^*|\mu)$. Our claim follows.

Step 3: Consequence of (17), convergence of the marginals.

As a consequence of Step 2, both sums in the r.h.s. of (17) are bounded sums of nonnegative terms. Therefore, they converge as $n \to +\infty$, and their terms tend to 0 as $k \to +\infty$. We deduce in particular that

$$H(\mu^*|\mu^{Q^n}) \underset{n\to+\infty}{\longrightarrow} 0.$$

In particular, by continuity of H w.r.t. its second variable as stated in Proposition 4, and by compactness of $\{\bar{\mu} \in \mathcal{M}_+(\mathcal{D}) \text{ s.t. } \mathsf{M}(\bar{\mu}) = \mathsf{M}(\nu)\}$, $\mu^{Q^n} \to \mu^*$. So now let us pick $\bar{Q}$ any limit point of (Q^n). Such a limit point exist by compactness of $\{Q \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F}) \text{ s.t. } \mathsf{M}(Q) = \mathsf{M}(\nu)\}$. It follows from $\mu^{Q^n} \to \mu^*$ that $\mu^{\bar{Q}} = \mu^*$.

Step 4: $\bar{Q} = Q^*$.

Let us show that $\bar{Q} = Q^*$, so that actually the whole sequence (Q^n) converges towards Q^*. On the one hand, passing to the limit $n \to +\infty$ along the subsequences generating $\bar{Q}$ in (17) and using the continuity of H w.r.t. the second variable as stated in Proposition 4, we find

$$H(Q|R) = H(Q|\bar{Q}) + \sum_{k=1}^{+\infty} \Big\{ H(\nu|\nu^{P^k}) - H(\mu^*|\mu) \Big\} + \sum_{k=1}^{+\infty} H(\mu^*|\mu^{Q^{k-1}}). \tag{18}$$

On the other hand, as for all $n \in \mathbb{N}^*$, $Q^n \ll R$, this is also true for $\bar{Q}$. In particular, $H(\bar{Q}|R) < +\infty$, and as $\bar{Q} \in \Pi(\mu^*, \nu)$, we can apply (18) with $\bar{Q}$ in place of Q, and find

$$H(\bar{Q}|R) = \sum_{k=1}^{+\infty} \Big\{ H(\nu|\nu^{P^k}) - H(\mu^*|\mu) \Big\} + \sum_{k=1}^{+\infty} H(\mu^*|\mu^{Q^{k-1}}). \tag{19}$$

Now it remains to apply (18) with $Q = Q^*$ and to plug the previous equality to find

$$H(Q^*|R) = H(Q^*|\bar{Q}) + H(\bar{Q}|R).$$

As by optimality of R^*, $H(\bar{Q}|R) \geq H(Q^*|R)$, we can conclude that $H(Q^*|Q) = 0$. Therefore, $\bar{Q} = Q^*$, as announced.

The proof of $P^n \to P^*$ follows the same lines. □

As a free output of the proof of Theorem 1, we can show that we could have swapped μ and $\bar{\mu}$, and ν and $\bar{\nu}$ in the definitions (13) of μ^* and ν^* respectively. This is justified in the following remark.

Remark 13. Observe the following optimization problem, where R, μ and ν are given, and where the competitor is $\bar{\nu}$:

$$\min \Big\{ H(\nu|\bar{\nu}) \,\Big|\, \bar{\nu} = \nu^P, \text{ for some } P \text{ with } H(P|R) < +\infty \text{ and } \mu^P = \mu \Big\}. \tag{20}$$

This problem is almost the same as the one defining ν^* in (13), except from the fact that ν and $\bar{\nu}$ are swapped in the relative entropy. In this remark, we justify that the solution of this problem is ν^* as well, and that the corresponding optimal value is $H(\mu^*|\mu)$.

Provided there exists a competitor $\bar{\nu}$ for this problem with $H(\nu|\bar{\nu}) < +\infty$, we can find P such that $H(P|R) < +\infty$ and $P \in \Pi(\mu, \bar{\nu})$, and $Q \in \mathcal{P}(\mathcal{D} \times \mathcal{F})$ defined for all i, j by

$$Q_{ij} := \frac{\nu_j}{\bar{\nu}_j} P_{ij},$$

which is legitimate since $H(\nu|\bar{\nu}) < +\infty$. We have then $H(Q|R) < +\infty$ and $\nu^Q = \nu$. Hence, using the definition (13) of μ^*, we have

$$H(\nu|\bar{\nu}) = H(Q|P) \geq H(\mu^Q|\mu) \geq H(\mu^*|\mu),$$

where the first equality is a direct computation, and where the first inequality is obtained using Corollary 6.

On the other hand, as soon as the assumption of Theorem 11 holds, ν^* is a competitor for the problem in (20), and so in particular $H(\nu|\nu^*) \geq H(\mu^*|\mu)$. But because the terms of the first series in (18) tend to 0 and $\nu^{P^k} \to \nu^*$, we conclude that actually, $H(\nu|\nu^*) = H(\mu^*|\mu)$ and ν^* is a solution of (20). Finally, it is easy to see that a solution $\bar{\nu}$ of (20) must satisfy $\bar{\nu} \ll \nu$ (because conditioning on the support of ν reduces the entropy), and by strict convexity of $\bar{\nu} \mapsto H(\nu|\bar{\nu})$ on the set $\{\bar{\nu} \ll \nu\}$, under the assumption of Theorem 11, the problem (20) admits ν^* as its unique solution, so that (20) can be used as an alternative definition of ν^*.

Of course, we could argue in the same way to provide an alternative definition of μ^*, and we have the following equalities:

$$H(\nu|\nu^*) = H(\mu^*|\mu) \qquad \text{and} \qquad H(\mu|\mu^*) = H(\nu^*|\nu).$$

In particular, $\mu^* \sim \mu$ and $\nu^* \sim \nu$ in the sense of measures.

We also give another remark concerning the generalization of Theorem 11 to Polish spaces.

Remark 14. We crucially use the fact that $\mathcal{D}$ and $\mathcal{F}$ are finite in order to obtain (18) and (19). In the continuous case, as H is not more than lower semicontinuous w.r.t the second variable, identity (18) becomes an inequality, where $=$ is replaced by $\geq$, which is the good direction for the proof. The difficulty is then to find an equality sign in (19).

4 Γ-convergence in the marginal penalization problem

In this section, we want to show that when R, μ and ν are such that the Schrödinger problem $\mathrm{Sch}(R;\mu,\nu)$ has no solution, then the limit points P^* and Q^* given by Theorem 11 are relevant in view of the possible applications of the Sinkhorn algorithm.

To do so, let us think of R as an imperfect theoretical model describing the coupling between the initial and final positions of the particles of a large system. Also, let us imagine that μ and ν are data obtained by measuring the positions of the particles of the actual system that R is supposed to describe, at the initial and final time. In this situation, if $\mathrm{Sch}(R;\mu,\nu)$ has a solution R^*, this solution is interpreted as the model that is the closest to R that can explain the data.

However, even when R is a rather good model, and when μ and ν are rather precise measurements, it is possible that $\mathrm{Sch}(R;\mu,\nu)$ has no solution for several reasons:

- The first reason could be that our modeling does not take into account some physical phenomena. For instance, in Subsection 4.1, we will consider the case where the true system allows creation or annihilation of mass with very small probability, whereas the modeling does not.
- Another reason could be that μ and ν are only approximations of the real marginals. This can result from imprecise or biased measurements, or from a restricted amount of collected data. This will be considered in Subsection 4.2.

In both cases, it is very natural to relax the marginal constraints in (10) by introducing a fitting term in the value functional, that cancels when the constraints are satisfied, but which remains finite otherwise.

The main result of this section asserts that in these two situations, that are actually very close, the limit points P^* and Q^* of the Sinkhorn algorithm allow to compute the solution of the relaxed problem when the new fitting term takes the form of an entropy, in the limit where the level of marginal penalization tends to $+\infty$. The second case is a direct consequence of the first one, but that we wanted to keep separated because it does not have the same interpretation.

4.1 Unbalanced problems

In this subsection, we give ourselves $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ as before, and we study the following optimization problem, which is a reasonable modification of $\mathrm{Sch}(R; \mu, \nu)$ where the marginal constraints are replaced with marginal penalizations:

$$\min \Big\{ H(\bar{R}|R) + \lambda\big(H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu)\big) \,\Big|\, \bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F}) \Big\}, \tag{21}$$

where $\lambda > 0$ parametrizes the level of penalization.

This approach is extremely reminiscent of the idea introduced by Liero, Mielke and Savaré in [20] to deal with unbalanced data, that is, when $\mathsf{M}(\mu) \neq \mathsf{M}(\nu)$, in optimal transport problems. This was the starting point of the theory of *unbalanced optimal transport*, also discovered independently by other teams [15, 6].

More precisely, we will study the limit of the problem in (21) as $\lambda \to +\infty$. In this limit, it is actually more convenient to call $\varepsilon = 1/\lambda$ and to multiply the value functional by ε, to find the problem that we call $\mathrm{Sch}^\varepsilon(R; \mu, \nu)$:

$$\mathrm{Sch}^\varepsilon(R; \mu, \nu) := \min \Big\{ \varepsilon H(\bar{R}|R) + H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu) \,\Big|\, \bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F}) \Big\}.$$

As we want to study the behavior of this problem in the limit $\varepsilon \to 0$, we define the following functionals:

$$\Lambda^\varepsilon : \bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F}) \mapsto \varepsilon H(\bar{R}|R) + H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu),$$
$$\Lambda : \bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F}) \mapsto \chi_{H(\bar{R}|R)<+\infty} + H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu),$$

where $\chi_{H(\bar{R}|R)<+\infty}$ is the convex indicatrix taking value 0 on the set

$$\Big\{ \bar{R} \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F}) \text{ such that } H(\bar{R}|R) < +\infty \Big\},$$

and $+\infty$ elsewhere.

The following proposition follows from standard arguments in the theory of Γ-convergence, see for instance [3, Theorem 1.47], and from the strict convexity of the relative entropy w.r.t. its first variable. We omit the proof.

Proposition 15. *We have:*

$$\Gamma - \lim_{\varepsilon \to 0} \Lambda^\varepsilon = \Lambda.$$

In particular, assuming that Λ is not uniformly infinite, let us call R_{opt} one of its minimizers, $\mu^g := \mu^{R_{\mathrm{opt}}}$ and $\nu^g := \nu^{R_{\mathrm{opt}}}$. The marginals μ^g and ν^g do not depend on the choice of R_{opt}, and as $\varepsilon \to 0$, the unique solution R^ε of $\mathrm{Sch}^\varepsilon(R; \mu, \nu)$ exists and converges towards the solution of $\mathrm{Sch}(R; \mu^g, \nu^g)$.

Remark 16. In the notations μ^g and ν^g, the g stands for *geometric*. This is because as shown in Theorem 17, μ^g and ν^g are respectively the componentwise geometric means of μ and μ^*, and of ν and ν^*.

Therefore, studying the behavior of $\mathrm{Sch}^\varepsilon(R;\mu,\nu)$ in the limit $\varepsilon \to 0$ reduces to the study of the Schrödinger problem with modified marginals μ^g and ν^g. The following theorem shows the link between R^* – the solution of $\mathrm{Sch}(R;\mu^g,\nu^g)$ – on the one hand, and P^* and Q^* from Theorem 11 on the other hand.

Theorem 17. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfy Assumption 8. Then the functional Λ is not uniformly infinite. Moreover, considering P^* and Q^* as given by Theorem 11, and μ^g and ν^g as given by Proposition 15, the solution of $\mathrm{Sch}(R;\mu^g,\nu^g)$ is the componentwise geometric mean of P^* and Q^*, that is, the matrix R^* defined for all i,j by*

$$R^*_{ij} := \sqrt{P^*_{ij}Q^*_{ij}}. \tag{22}$$

Also, if μ^* and ν^* are defined by (13), μ^g and ν^g are the componentwise geometric means of μ^* and μ for the first one, and of ν^* and ν for the second one. In other terms, we have for all i,j,

$$\mu^g_i = \sqrt{\mu^*_i\mu_i} \qquad \text{and} \qquad \nu^g_j = \sqrt{\nu^*_j\nu_j}. \tag{23}$$

Remark 18.

- Having in mind the approach of [20], we can give the following interpretation of the matrix R^*: In the degenerate case where the Schrödinger problem has no solution, it is necessary to allow creation and annihilation of mass to find solutions. Following [20], we can do this by replacing the balanced problem $\mathrm{Sch}(R;\mu,\nu)$ by the unbalanced problem $\mathrm{Sch}^\varepsilon(R;\mu,\nu)$. Following this analogy, $\lambda = \frac{1}{\varepsilon}$ parametrizes the cost of creating particles. The matrix R^* from Theorem 17 is therefore the limit of these solutions when the cost of creating or destroying matter tends to $+\infty$.

- A small adaptation of the proof shows that given $\alpha \in [0,1]$, if we replace the problem in (21) by

$$\min\Big\{H(\bar{R}|R) + \lambda\Big((1-\alpha)H(\mu^{\bar{R}}|\mu) + \alpha H(\nu^{\bar{R}}|\nu)\Big)\ \Big|\ \bar{R}\in\mathcal{M}_+(\mathcal{D}\times\mathcal{F})\Big\},$$

  and if we call $R^{\alpha,\lambda}$ its solution, then as $\lambda\to+\infty$, we have for all i,j:

$$R^{\alpha,\lambda}_{ij} \underset{\lambda\to+\infty}{\longrightarrow} \big(P^*_{ij}\big)^{1-\alpha}\big(Q^*_{ij}\big)^{\alpha}.$$

To prove this theorem, we will need to study carefully the optimality conditions for μ^* and ν^*. This could be done writing the Karush-Kuhn-Tucker conditions for the corresponding optimalization problems. We will rather adopt a more hand by hand approach, that is more likely to be generalizable in the continuous case. This is done in the following proposition.

Proposition 19. *Assume that the conditions of Theorem 11 are fulfilled. For all i,j, we have*

$$P^*_{ij} = \frac{\mu_i}{\mu^*_i}Q^*_{ij} \qquad \text{and} \qquad Q^*_{ij} = \frac{\nu_j}{\nu^*_j}P^*_{ij}, \tag{24}$$

with convention $\frac{0}{0}=0$. In particular, P^* and Q^* are equivalent, and we call $\mathcal{S}$ their common support. Also, recall the definition of $\mathcal{E}$ in (11). Of course $\mathcal{S}\subset\mathcal{E}$. Finally, we call for all i,j

$$\varphi_i := \log\frac{\mu^*_i}{\mu_i} \qquad \text{and} \qquad \psi_j := \log\frac{\nu^*_j}{\nu_j}. \tag{25}$$

For all $(i,j)\in\mathcal{E}$, φ_i and ψ_j are well defined in $\mathbb{R}$, and:

$$\begin{cases} \varphi_i+\psi_j = 0, & \text{if } (i,j)\in\mathcal{S},\\ \varphi_i+\psi_j \geq 0, & \text{if } (i,j)\in\mathcal{E}. \end{cases} \tag{26}$$

Proof of Proposition 19. To get (24), it suffices to let n tend to $+\infty$ in (15). The fact that $\mathcal{S} \subset \mathcal{E}$ relies on the closed property of P^n and Q^n defined in (1) to have its support included in $\mathcal{E}$ for $n \geq 2$. If $(i,j) \in \mathcal{E}$, let us check that φ_i and ψ_j are well defined. On the one hand, by definition of $\mathcal{E}$, i is in the support of μ and j is in the support of ν. On the other hand, as observed in Remark 13, $\mu^* \sim \mu$ and $\nu^* \sim \nu$. Our claim follows.

Now, let $(i,j) \in \mathcal{S}$. A consequence of (24) is

$$P^*_{ij} = \frac{\mu^*_i}{\mu_i}\frac{\nu^*_j}{\nu_j}P^*_{ij} = \exp(\varphi_i + \psi_j)P^*_{ij}.$$

As (i,j) is in the support of P^* by definition of $\mathcal{S}$, we conclude that $\varphi_i + \psi_j = 0$.

Finally, it remains to prove that for all $(i,j) \in \mathcal{E}$, $\varphi_i + \psi_j \geq 0$. For this we use the optimality of $H(\mu^*|\mu) = H(\mu^{Q^*}|\mu)$ over all Q satisfying $\nu^Q = \nu$. So let us take $(i,j) \in \mathcal{E}$. As $\nu_j > 0$, there exists i' such that $(i',j) \in \mathcal{S}$, that is, such that $Q^*_{i'j} > 0$. Let us define for $\varepsilon > 0$

$$Q^\varepsilon = Q^* + \varepsilon\delta_{ij} - \varepsilon\delta_{i'j},$$

where δ_{ij} is the matrix whose only nonzero coefficient is a one at position (i,j), and similarly for $\delta_{i'j}$. If ε is sufficiently small, $Q^\varepsilon \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\nu^{Q^\varepsilon} = \nu$ and with obvious notations, $\mu^{Q^\varepsilon} = \mu^* + \varepsilon\delta_i - \varepsilon\delta_{i'}$. Therefore, for such ε,

$$H(\mu^{Q^\varepsilon}|\mu) \geq H(\mu^*|\mu).$$

derivating to the right this inequality at $\varepsilon = 0$, we find

$$\log\frac{\mu^*_i}{\mu_i} - \log\frac{\mu^*_{i'}}{\mu_{i'}} \geq 0,$$

which rewrites $\varphi_i - \varphi_{i'} \geq 0$. But $(i',j) \in \mathcal{S}$ so $\varphi_{i'} = -\psi_j$, and so $\varphi_i + \psi_j \geq 0$. □

With this proposition at hand, we can prove Theorem 17.

Proof of Theorem 17. The fact that under Assumption 8, Λ is not uniformly infinite follows from observing that $\Lambda(R^0) < +\infty$, where R_0 was defined Assumption 8. Now we reason in two steps. First we will prove using Proposition 19 that R^* defined by (22) is an optimizer of Λ, and then that it is the solution of the Schrödinger problem between its marginals.

Step 1: R^* is an optimizer of Λ.

To see that R^* is an optimizer of Λ, we first give a formula relating the vectors φ and ψ as defined by formula (25) and the marginals μ^{R^*} and ν^{R^*} of R^*. Using (24) and the definition (22) of R^*, we see that for all i,j,

$$R^*_{ij} = \sqrt{\frac{\nu_j}{\nu^*_j}}P^*_{ij} = \sqrt{\frac{\mu_i}{\mu^*_i}}Q^*_{ij}. \tag{27}$$

Summing respectively these identities w.r.t. i and j, we deduce that for all i,j,

$$\mu^{R^*}_i = \sqrt{\mu^*_i\mu_i} = \sqrt{\frac{\mu^*_i}{\mu_i}}\mu_i \qquad \text{and} \qquad \nu^{R^*}_j = \sqrt{\nu^*_j\nu_j} = \sqrt{\frac{\nu^*_j}{\nu_j}}\nu_j.$$

Let us define for all i,j:

$$Z^\mu_i := \log\frac{\mu^{R^*}_i}{\mu_i} = \frac{1}{2}\varphi_i \qquad \text{and} \qquad Z^\nu_j := \log\frac{\nu^{R^*}_j}{\nu_j} = \frac{1}{2}\psi_j.$$

Note that for all $(x_i,y_j) \in \mathcal{E}$, Z^μ_i and Z^ν_j are well defined in $\mathbb{R}$.

Now let $\bar{R}$ be such that $\Lambda(\bar{R}) < +\infty$. Using inequality (5) to bound from below each relative entropy, we have

$$\begin{aligned}\Lambda(\bar{R}) &= H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu)\\ &\geq \langle Z^\mu, \mu^{\bar{R}}\rangle - \langle e^{Z^\mu} - 1, \mu\rangle + \langle Z^\nu, \nu^{\bar{R}}\rangle - \langle e^{Z^\nu} - 1, \nu\rangle\\ &= \frac{1}{2}\langle \varphi, \mu^{\bar{R}}\rangle + \frac{1}{2}\langle \psi, \nu^{\bar{R}}\rangle - \sum_i \left\{\mu_i^{R^*} - \mu_i\right\} - \sum_j \left\{\nu_j^{R^*} - \nu_j\right\}\\ &= \frac{1}{2}\langle \varphi \oplus \psi, \bar{R}\rangle + \mathsf{M}(\mu) + \mathsf{M}(\nu) - 2\mathsf{M}(R^*),\end{aligned}$$

where $\varphi\oplus\psi$ is the matrix defined for all i, j by $(\varphi\oplus\psi)_{ij} := \varphi_i + \psi_j$. Now, because of the second line of (26), as the support of $\bar{R}$ is easily seen to be a subset of $\mathcal{E}$, we get

$$\Lambda(\bar{R}) \geq \mathsf{M}(\mu) + \mathsf{M}(\nu) - 2\mathsf{M}(R^*).$$

On the other hand, by definition of Z^μ and Z^ν,

$$\begin{aligned}\Lambda(R^*) &= H(\mu^{R^*}|\mu) + H(\nu^{R^*}|\nu)\\ &= \langle Z^\mu, \mu^{R^*}\rangle + \mathsf{M}(\mu) - \mathsf{M}(R^*) + \langle Z^\nu, \nu^{R^*}\rangle + \mathsf{M}(\nu) - \mathsf{M}(R^*)\\ &= \frac{1}{2}\langle \varphi \oplus \psi, R^*\rangle + \mathsf{M}(\mu) + \mathsf{M}(\nu) - 2\mathsf{M}(R^*).\end{aligned}$$

But now, as the support of R^* is precisely $\mathcal{S}$, by the first line of (26), we get

$$\Lambda(R^*) = \mathsf{M}(\mu) + \mathsf{M}(\nu) - 2\mathsf{M}(R^*).$$

We deduce that $\Lambda(\bar{R}) \geq \Lambda(R^*)$ and R^* is indeed an optimizer of Λ. In particular, $\mu^g = \mu^{R^*}$ and $\nu^g = \nu^{R^*}$, which proves (23).

Step 2: R^* is the solution of $\mathrm{Sch}(R; \mu^g, \nu^g)$.

To show that R^* solves the Schrödinger problem between its marginals, we consider another $\bar{R} \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$ such that $\bar{R} \in \Pi(\mu^g, \nu^g)$ and $H(\bar{R}|R) < +\infty$. Then, for $\varepsilon > 0$, we define

$$P^\varepsilon := P^* + \varepsilon(\bar{R} - R^*).$$

As $R^* \ll P^*$ (see (22)), whenever ε is sufficiently small, $P^\varepsilon \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, and in addition, we easily check that $P^\varepsilon \in \Pi(\mu, \nu^*)$. So by definition (14) of P^*,

$$H(P^*|R) \leq H(P^\varepsilon|R).$$

Derivating this inequality to the right at $\varepsilon = 0$, we find

$$\sum_{ij} R^*_{ij} \log \frac{P^*_{ij}}{R_{ij}} \leq \sum_{ij} \bar{R}_{ij} \log \frac{P^*_{ij}}{R_{ij}},$$

with convention $\frac{0}{0} = 0$, $0\log 0 = 0$ and $a\log 0 = -\infty$ for all $a > 0$. In particular, we deduce that $\bar{R} \ll P^* \sim R^*$, and our inequality rewrites

$$H(R^*|R) + H(\bar{R}|P^*) - H(R^*|P^*) \leq H(\bar{R}|R).$$

The last thing to observe is that because of (27), R^* is the solution of the Schrödinger problem $\mathrm{Sch}(P^*; \mu^g, \nu^g)$: a direct application of (5) with $Z_{ij} = \log \frac{R^*_{ij}}{P^*_{ij}} = -\bar{\psi}_j$ (which is well defined on the support of P^*, and so on the support of $\bar{R}$) provides

$$\begin{aligned}H(\bar{R}|P^*) &\geq \langle Z, \bar{R}\rangle - \langle e^Z - 1, P^*\rangle\\ &= -\langle \bar{\psi}, \nu^g\rangle + \sum_{ij} P^*_{ij} - R^*_{ij}\\ &= \langle Z, R^*\rangle + \sum_{ij} P^*_{ij} - R^*_{ij} = H(R^*|P^*).\end{aligned}$$

The result follows. □

Remark 20. In Step 2, we used a particular case of the following more general result that is proved in the same way:

Lemma 21. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu, \mu' \in \mathcal{M}_+(\mathcal{D})$ and $\nu, \nu' \in \mathcal{M}_+(\mathcal{F})$. Assume that $\mathrm{Sch}(R;\mu,\nu)$ admits a solution P and that $\mathrm{Sch}(P;\mu',\nu')$ admits a solution Q. Then the unique solution of $\mathrm{Sch}(R;\mu',\nu')$ exists: it is Q.*

4.2 Balanced version

In the last subsection, we interpreted the fact that $\mathrm{Sch}(R;\mu,\nu)$ has no solution by the fact that our model does not incorporate the ability of the real system to create or destroy mass. In that case, the total mass of R^* is not the same as the one of μ and ν in general, even when the latter two coincide. Therefore, R^* cannot be interpreted directly as a joint law for the initial and final positions of the particles. Following the lines of [20], we see that its interpretation is actually rather complicated.

In this subsection, we want to consider the case where the real system under study is truly balanced, that is, no creation of annihilation of mass is possible at all. In this situation, whatever the way we are obtaining the data, μ and ν must have the same mass, and up to renormalizing, we can assume that they are probability measures. We want to interpret the fact that $\mathrm{Sch}(R;\mu,\nu)$ has no solution by the fact that μ and ν are imperfect measurements of the true marginals, and we want to find a *probability* measure $\bar{R}^*$ that is entropically close to R while having its marginals entropically close to μ and ν, that can be interpreted as a joint law.

Therefore, we introduce the following problem that is a slight modification of Sch^ε where the competitor $\bar{R}$ needs to be a probability measure: for all $R \in \mathcal{P}(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{P}(\mathcal{D})$ and $\nu \in \mathcal{P}(\mathcal{F})$,

$$\overline{\mathrm{Sch}}^\varepsilon(R;\mu,\nu) := \min\left\{\varepsilon H(\bar{R}|R) + H(\mu^{\bar{R}}|\mu) + H(\nu^{\bar{R}}|\nu) \,\middle|\, \bar{R} \in \mathcal{P}(\mathcal{D}\times\mathcal{F})\right\}.$$

The following theorem states the behaviour of this optimization problem as $\varepsilon \to 0$, and is a direct adaptation of Theorem 17 to the balanced case.

Theorem 22. *Let $R \in \mathcal{P}(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{P}(\mathcal{D})$ and $\nu \in \mathcal{P}(\mathcal{F})$ satisfy the conditions of Assumption 8, and call*

$$\mathcal{Z} := \sum_{ij} \sqrt{P^*_{ij} Q^*_{ij}},$$

where P^* and Q^* are given by Theorem 11. Then for all $\varepsilon > 0$, the solution $\bar{R}^\varepsilon$ of $\overline{\mathrm{Sch}}^\varepsilon(R;\mu,\nu)$ exists, is unique, and satisfies for all i,j:

$$\bar{R}^\varepsilon_{ij} \underset{\varepsilon\to 0}{\longrightarrow} \bar{R}^*_{ij} := \frac{\sqrt{P^*_{ij}Q^*_{ij}}}{\mathcal{Z}}.$$

Its marginals are given for all i,j by

$$\mu_i^{\bar{R}^*} = \frac{\sqrt{\mu_i^*\mu_i}}{\mathcal{Z}} \qquad \text{and} \qquad \nu_j^{\bar{R}^*} = \frac{\sqrt{\nu_j^*\nu_j}}{\mathcal{Z}}.$$

Proof. Theorem 22 is a direct consequence of Theorem 17 once noticed the following fact: If R, μ, ν are as in the statement of the theorem, if $\varepsilon > 0$ and if R^ε is the solution of $\mathrm{Sch}^\varepsilon(R;\mu,\nu)$, then $R^\varepsilon/\mathsf{M}(R^\varepsilon)$ is the solution of $\overline{\mathrm{Sch}}^\varepsilon(R;\mu,\nu)$. To see this, consider $R' \in \mathcal{P}(\mathcal{D}\times\mathcal{F})$. Direct computations imply

$$H(R'|R) = \frac{H\left(\mathsf{M}(R^\varepsilon)R' \middle| R\right)}{\mathsf{M}(R^\varepsilon)} + \log\frac{1}{\mathsf{M}(R^\varepsilon)} + 1 - \frac{1}{\mathsf{M}(R^\varepsilon)},$$
$$H\left(\frac{R^\varepsilon}{\mathsf{M}(R^\varepsilon)} \middle| R\right) = \frac{H(R^\varepsilon|R)}{\mathsf{M}(R^\varepsilon)} + \log\frac{1}{\mathsf{M}(R^\varepsilon)} + 1 - \frac{1}{\mathsf{M}(R^\varepsilon)}.$$

By optimality of R^ε, $H(\mathsf{M}(R^\varepsilon)R'|R) \geq H(R^\varepsilon|R)$, and therefore $H(R'|R) \geq H(R^\varepsilon/\mathsf{M}(R^\varepsilon)|R)$. Our claims follows, and hence the theorem as M is a continuous functional and $\mathcal{Z} = \mathsf{M}(R^*)$, where R^* is given by Theorem 17. □

5 Existence and support of the solutions to Schrödinger problems

In this section, our goal is to give a detailed study of the support of the solution of $\mathrm{Sch}(R;\mu,\nu)$ when the latter exists, or of the common one of P^*, Q^* and R^* from Theorems 11 and 17 in the non-scalable case. This study will rely on a new interpretation of the well known existence conditions for the Schrödinger problem in finite spaces, for which we refer to [4, 14].

We start with our new formulation of these conditions of existence, which is very close to the ones introduced by Brualdi [4], but has the advantage of helping understanding the shape of the support of the optimizers seen as a bipartite graph.

In the second part of the section, we provide a theoretical procedure allowing to get the support of the optimizers, both in the approximately scalable and non-scalable cases, without using the Sinkhorn algorithm. This procedure will be used in the next section as a preliminary step, before launching the Sinkhorn algorithm, in order to recover a linear rate for the latter.

5.1 A necessary and sufficient condition of existence for the Schrödinger problem in finite spaces

Let us state a necessary and sufficient condition on R, μ and ν for the existence of a solution R^* of $\mathrm{Sch}(R;\mu,\nu)$, that is, for $\mathrm{Sch}(R;\mu,\nu)$ to be scalable or approximately scalable. In order to do so, we need to give a few definitions. First, we endow the set $\mathcal{D}\cup\mathcal{F}$ with a bipartite graph structure related to R: we set

$$\forall i = 1,\dots,N \text{ and } j = 1,\dots,M, \qquad x_i \triangle y_j \Leftrightarrow R_{ij} > 0.$$

We have $x_i\triangle y_j$ whenever it is possible to travel from x_i to y_j under R. We write indifferently $x_i\triangle y_j$ or $y_j\triangle x_i$.

With this structure in hand, we are able to push forward or pull backward subsets of $\mathcal{D}$ and $\mathcal{F}$, that is, we define:

$$\begin{aligned} \forall A \subset \mathcal{D}, \quad F_R(A) &:= \Big\{ y \in \mathcal{F} \,|\, \exists x \in A \text{ s.t. } x\triangle y \Big\}, \\ \forall B \subset \mathcal{F}, \quad D_R(B) &:= \Big\{ x \in \mathcal{D} \,|\, \exists y \in B \text{ s.t. } x\triangle y \Big\}. \end{aligned} \tag{28}$$

Heuristically, for all $A \subset \mathcal{D}$, $F_R(A)$ is the set of all possible final positions of particles starting from A, under R. Correspondingly, for all $B \subset \mathcal{F}$, $D_R(B)$ is the set of all possible initial positions of particles arriving in B under R. Notice the explicit mention of R in the notations: in the following, we will allow ourselves to replace R by any other measure $\bar{R} \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$.

The main result of this section is the following.

Theorem 23. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$. The three following assertions are equivalent:*

(a) $\mathsf{M}(\mu) = \mathsf{M}(\nu)$ *and for all* $A \subset \mathcal{D}$, $\mu(A) \leq \nu(F_R(A))$.

(b) $\mathsf{M}(\mu) = \mathsf{M}(\nu)$ *and for all* $B \subset \mathcal{F}$, $\nu(B) \leq \mu(D_R(B))$.

(c) $\mathrm{Sch}(R;\mu,\nu)$ *is scalable or approximately scalable.*

Note that the implications (c) ⇒ (a) and (c) ⇒ (b) are straightforward, and that only the reverse implications are challenging. Also, we already noticed in Subsection 2.3 that (c) implies Assumption 8. Hence, it is also the case for (a) and (b).

The proof relies on the following Lemma 24, which gives a necessary and sufficient condition on R, μ and ν ensuring R^* to have the same support as R, that is, to be in the scalable case. In this statement, we use the notations μ^R and ν^R as defined in (3), and we work under Assumption 9, which is always possible under Assumption 8 up to considering subspaces of $\mathcal{D}$ and $\mathcal{F}$, see Subsection 2.3.

Lemma 24. *Let $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$, satisfying Assumption 9. The three following assertions are equivalent:*

(a') $\mathsf{M}(\mu) = \mathsf{M}(\nu)$ *and for all $A \subset \mathcal{D}$, $\mu(A) \leq \nu(F_R(A))$, with a strict inequality whenever $\mu^R(A) < \nu^R(F_R(A))$.*

(b') $\mathsf{M}(\mu) = \mathsf{M}(\nu)$ *and for all $B \subset \mathcal{F}$, $\nu(B) \leq \mu(D_R(B))$, with a strict inequality whenever $\nu^R(B) < \mu^R(D_R(B))$.*

(c') $\mathrm{Sch}(R; \mu, \nu)$ *is scalable.*

In plain words, it highlights the difference between the approximately scalable and scalable cases, by showing that the scalable case consists in assuming as much strict inequalities in (a) or in (b) as possible. Although both Theorem 23 and Lemma 24 can be directly deduced from the work of Brualdi [4], we provide in Appendix B a short and independent proof based on topological arguments.

5.2 Theoretical construction of the support

In the scalable case, the Sinkhorn algorithm is known to have a linear rate of convergence. On the other hand, in the approximately scalable case, the algorithm still converges, but the (unknown) convergence rate cannot be linear [1].

In this subsection, we study the support of the solution of the Schrödinger problem in the approximately scalable and non-scalable cases for the following reason. Take R, μ and ν such that $\mathrm{Sch}(R; \mu, \nu)$ is approximately scalable, R^* the solution of this problem, and $\mathcal{S}$ the support of R^*. Then $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}} R; \mu, \nu)$ is scalable and its solution is R^*. In particular, the Sinkhorn algorithm applied to this problem has a linear rate of convergence. Interestingly, a similar reasoning is valid in the non-scalable case, as we show in Proposition 25 below.

Without loss of generality, and for the sake of simplicity, in the whole subsection, we work under Assumption 9. By Remark 13, if μ^* and ν^* are defined by (13), we have $\nu \sim \nu^*$ and $\mu \sim \mu^*$. So under Assumption 9, they have a full support as well.

Proposition 25. *Let $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9. Let us call $\mathcal{S}$ the common support of P^* and Q^* from Theorem 11, and R^* from Theorem 17. Let $(P^n)_{n \in \mathbb{N}^*}$ and $(Q^n)_{n \in \mathbb{N}^*}$ be given by (2) applied to $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}} R; \mu, \nu)$. They converge respectively towards P^* and Q^*, both of them at a linear rate.*

Proof. Let $(P^n)_{n \in \mathbb{N}^*}$ and $(Q^n)_{n \in \mathbb{N}^*}$ be given by the equivalent formulations (1) and (2) applied to $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}} R; \mu, \nu)$. Let us show that (P^n) converges towards P^* at a linear rate. The case of (Q^n) follows the same arguments. The idea is that if $(\tilde{P}^n)_{n \in \mathbb{N}^*}$ and $(\tilde{Q}^n)_{n \in \mathbb{N}^*}$ are given by (1) and (2) applied to $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}} R; \mu, \nu^*)$, then for all $n \in \mathbb{N}^*$, $P^n = \tilde{P}^n$. As the problem $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}} R; \mu, \nu^*)$ is scalable (its solution, P^*, has the same support as $\mathbb{1}_{\mathcal{S}} R$), the rate of convergence of $(\tilde{P}^n)$ towards P^* is linear, and the result follows.

So let us prove by induction that for all $n \in \mathbb{N}^*$, $P_n = \tilde{P}_n$. According to (1), P^1 and $\tilde{P}^1$ are solutions to the same problem, and therefore coincide. Let us now consider $n \in \mathbb{N}^*$ such that

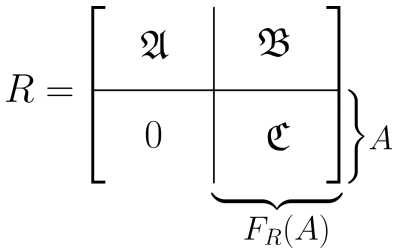


Figure 2 : If $A \subset \mathcal{D}$, up to reordering the lines, we can assume that it corresponds to the last lines. Up to reordering the columns, we can assume that $F_R(A)$ corresponds to the last columns. Then, R has the form given in the picture. In this situation, A is a SISP set for $(R;\mu,\nu)$ if R^* cancels on the block $\mathfrak{B}$ and if the supports of R and R^* coincide on the block $\mathfrak{C}$.

$P^n = \tilde{P}^n$ and show that $P^{n+1} = \tilde{P}^{n+1}$. By construction, the support of P^{n+1} and $\tilde{P}^{n+1}$ is $\mathcal{S}$, so we just need to check that for all $(x_j, y_j) \in \mathcal{S}$, $P^{n+1}_{ij} = \tilde{P}^{n+1}_{ij}$. By the first line of (26), for all $(x_i, y_j) \in \mathcal{S}$, we have:

$$\nu_j = \frac{\mu_i^* \nu_j^*}{\mu_i}.$$

Hence, for all $(x_i, y_j) \in \mathcal{S}$:

$$P^{n+1}_{ij} = \frac{\mu_i}{\mu_i^{Q^n}} Q^n_{ij} = \frac{\mu_i}{\sum_{j'} \frac{\nu_{j'}}{\nu_{j'}^{P^n}} P^n_{ij'}} \times \frac{\nu_j}{\nu_j^{P^n}} P^n_{ij} = \frac{\mu_i}{\frac{\mu_i^*}{\mu_i} \sum_{j'} \frac{\nu_{j'}^*}{\nu_{j'}^{\tilde{P}^n}} \tilde{P}^n_{ij'}} \times \frac{\mu_i^*}{\mu_i} \frac{\nu_j^*}{\nu_j^{\tilde{P}^n}} \tilde{P}^n_{ij} = \frac{\mu_i}{\mu_i^{\tilde{Q}^n}} \tilde{Q}^n_{ij} = \tilde{P}^{n+1}_{ij},$$

where the change from P^n to $\tilde{P}^n$ in the middle coming from the induction assumption $P^n = \tilde{P}^n$. The result follows. □

Therefore, even in the non-scalable case, a way to improve the Sinkhorn algorithm consists in first finding $\mathcal{S}$, and then computing the solution of a scalable problem. We propose in this subsection a theoretical procedure allowing to get this support without using the Sinkhorn algorithm in both the approximately and non-scalable cases, and we will propose an approximate method for achieving this task numerically at Section 6.

To detail our procedure, we introduce a class of subsets of $\mathcal{D}$ associated with a triple $(R;\mu,\nu)$.

Definition 26. Let $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9. Let us consider R^* from Theorem 17. We say that a subset $A \subset \mathcal{D}$ is the *source of an isolated scalable problem* (or for short that A is a SISP set) for $(R;\mu,\nu)$ if $A \neq \emptyset$ and:

- The set $(\mathcal{D}\backslash A) \times F_R(A)$ is R^*-negligible, *i.e.*

$$R^*\Big((\mathcal{D}\backslash A) \times F_R(A)\Big) = 0. \tag{29}$$

- For all $x_i \in A$ and $y_j \in \mathcal{F}$,

$$R^*_{ij} > 0 \quad \Leftrightarrow \quad R_{ij} > 0. \tag{30}$$

We show at Figure 2 an illustration of what a SISP is.

Of course, as P^* and Q^* from Theorem 11 are equivalent to R^* in the sense of measures, we could have replaced R^* in the previous definition by one of them.

On the one hand, SISP sets always exist, at least under Assumption 9, as announced in the following lemma. Its proof is our main task in this part of our work, and is given at the end of the subsection.

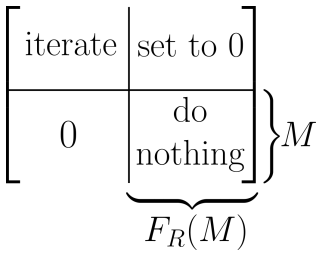


Figure 3 : In the situation of Figure 2 where we have reordered the lines and columns, our procedure consists in recursively add zeros to R at positions where we know thanks to (29) that R^* admits a zero. We know that we did not forget any zero in $M \times F_R(M)$ thanks to (35).

Lemma 27. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9. Then there exists a SISP set for $(R;\mu,\nu)$.*

On the other hand, once we know how to find SISP sets, an iterative procedure consisting in finding SISP sets for a sequence of more and more restricted problems makes is possible to reconstruct the whole subset $\mathcal{S}$.

Proposition 28. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9. Let us call $\mathcal{S}$ the common support of P^* and Q^* from Theorem 11, and R^* from Theorem 17.*

We define by inference $(R^n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $(\mathcal{D}^n)_{n\in\mathbb{N}}$ a nonincreasing sequence of subsets of $\mathcal{D}$ and $(\mathcal{F}^n)_{n\in\mathbb{N}}$ a nonincreasing sequence of subsets of $\mathcal{F}$ in the following way:

- *For $n=0$, we set $R^0 := R$, $\mathcal{D}^0 := \mathcal{D}$ and $\mathcal{F}^0 := \mathcal{F}$;*
- *For all $n \in \mathbb{N}$, if $\mathcal{D}^n$ and $\mathcal{F}^n$ are nonempty and $(R^n{}_{\llcorner \mathcal{D}^n\times\mathcal{F}^n}; \mu_{\llcorner\mathcal{D}^n}, \nu_{\llcorner\mathcal{F}^n})$ satisfies Assumption 9, we pick M_n a SISP set as given by Lemma 27, and we set:*

$$\mathcal{D}^{n+1} := \mathcal{D}^n \backslash M_n, \qquad \mathcal{F}^{n+1} := \mathcal{F}^n \backslash F_{R^n{}_{\llcorner\mathcal{D}^n\times\mathcal{F}^n}}(M_n),$$

$$\forall i,j, \quad R^{n+1}_{ij} := \begin{cases} 0, & \text{if } y_j \in F_{R^n{}_{\llcorner\mathcal{D}^n\times\mathcal{F}^n}}(M_n) \text{ and } x_i \in \mathcal{D}^{n+1}, \\ R^n_{ij}, & \text{otherwise.} \end{cases}$$

Otherwise, we set $R^{n+1} := R^n$, $\mathcal{D}^{n+1} := \mathcal{D}^n$ and $\mathcal{F}^{n+1} := \mathcal{F}^n$.

With this construction, the sequence $(R^n, \mathcal{D}^n, \mathcal{F}^n)_{n\in\mathbb{N}}$ is stationary. More precisely, there exists $N \in \mathbb{N}^*$ such that for all $n \geq N$,

$$\mathcal{D}^n = \emptyset, \qquad \mathcal{F}^n = \emptyset, \qquad R^n = \mathbb{1}_{\mathcal{S}} R.$$

An illustration of the procedure at each iteration, is provided in Figure 3. An illustration of the full procedure in a specific non-scalable case is provided in Figure 4.

Proof. In this proof, in order to lighten the notations, we call $R^n_r := R^n{}_{\llcorner\mathcal{D}^n\times\mathcal{F}^n}$. We will prove by inference the following facts. For all $n \in \mathbb{N}$:

1. Calling $\mathcal{S}^n$ the support of R^n, and therefore $\mathcal{S}^0$ the support of R,

$$\mathcal{S}^n \cap \Big((\mathcal{D}\times\mathcal{F})\backslash(\mathcal{D}^n\times\mathcal{F}^n)\Big) = \mathcal{S} \cap \Big((\mathcal{D}\times\mathcal{F})\backslash(\mathcal{D}^n\times\mathcal{F}^n)\Big) = \mathcal{S}^0 \cap \left(\bigcup_{k=0}^{n-1} M_k \times F_{R^k_r}(M_k)\right), \quad (31)$$

$\mathcal{S} \subset \mathcal{S}^n$, and $\mathcal{S}^n \cap (\mathcal{D}^n \times \mathcal{F}^n) = \mathcal{S}^0 \cap (\mathcal{D}^n \times \mathcal{F}^n)$.

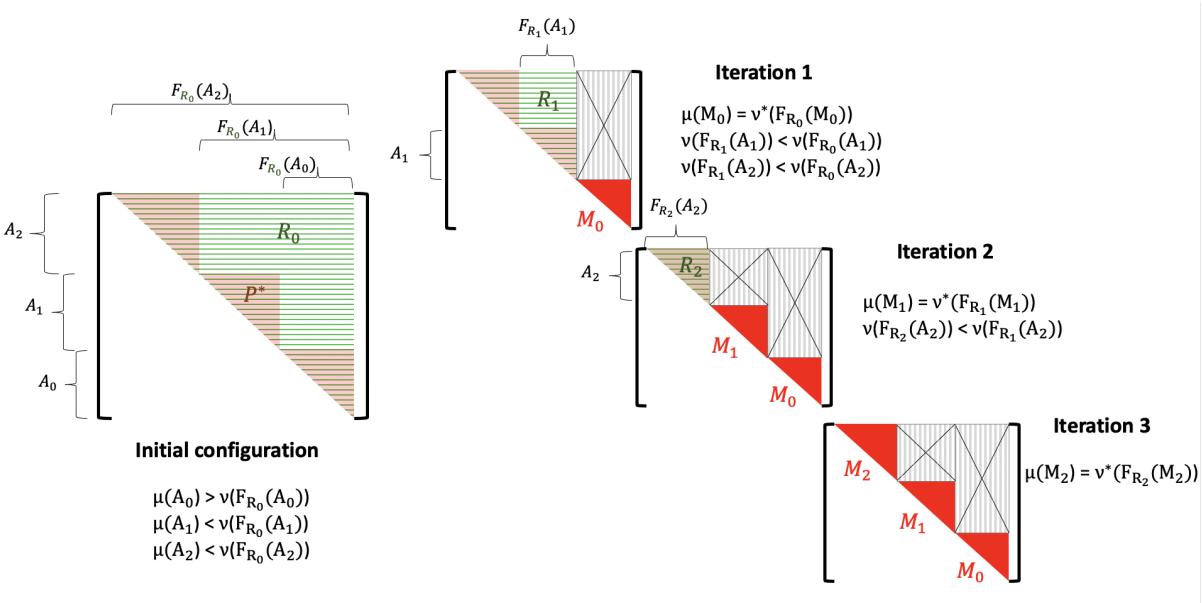


Figure 4 : Illustration of the procedure of Proposition 28 when the matrix R is upper diagonal and R^* a staircase matrix (see Appendix A for more details). Only A_0 is a SISP set for $(R;\mu,\nu)$, but A_1 and A_2 are SISP sets for the restricted problems at the iterations 2 and 3. In this example, the procedure is stationary after 3 steps. In this example, the SISP set at each iteration is the unique maximal θ-set existing for the reduced problem, as presented in Definition 30. We remark that we can also build ν^*, the second marginal of P^* defined in Theorem 11, on the successive SISP sets obtained along the procedure, thanks to the second step of the proof of Proposition 27 which ensures that the ratio $\frac{\nu}{\nu^*}$ is constant inside the maximal θ-sets.

2. $\mathcal{D}^n$ is empty if and only if $\mathcal{F}^n$ is empty.

3. If $\mathcal{D}^n$ and $\mathcal{F}^n$ are not empty, $(R_r^n;\mu_{\llcorner\mathcal{D}^n},\nu_{\llcorner\mathcal{F}^n})$ satisfies Assumption 9 and the matrices $P^{n,*}$, $Q^{n,*}$ and $R^{n,*}$ associated with $(R_r^n;\mu_{\llcorner\mathcal{D}^n},\nu_{\llcorner\mathcal{F}^n})$ through Theorems 11 and 17 are the restrictions of P^*, Q^* and R^* to $\mathcal{D}^n\times\mathcal{F}^n$.

This is enough to prove the proposition: if the conclusion of the inference is true, then by the third point and Lemma 27, as long as $\mathcal{D}^n$ and $\mathcal{F}^n$ are nonempty, $(R_r^n;\mu_{\llcorner\mathcal{D}^n},\nu_{\llcorner\mathcal{F}^n})$ admits a SISP set M_n, which is not empty by definition. Therefore, $(\mathcal{D}^n)$ is strictly decreasing in the sense of inclusion as long as it is not empty, so it has to reach $\emptyset$ at a certain rank N. At this rank, because of the first point, we also have $\mathcal{F}^N=\emptyset$, and because of (31), $\mathcal{S}^N=\mathcal{S}$, so that the conclusion follows. So let us prove the inference.

At rank 0, everything is clear, so let us assume that the conclusions of points one, two and three hold at rank n, and prove them at rank $n+1$. First, if $\mathcal{D}^n$ is empty, by assumption $\mathcal{F}^n$ is empty as well, so we have reached a stationary point, and everything is still true at rank $n+1$. So we can assume without loss of generality that $\mathcal{D}^n$, and hence $\mathcal{F}^n$, are nonempty. By assumption, $(R_r^n;\mu_{\llcorner\mathcal{D}^n},\nu_{\llcorner\mathcal{F}^n})$ satisfies Assumption 9, and by Lemma 27, we can find a SISP set M_n. In this context, let us check the points one by one at rank $n+1$.

First point. Observing that $\mathcal{D}^n$ is the disjoint union of M_n and $\mathcal{D}^{n+1}$, and that $\mathcal{F}^n$ is the disjoint union of $F_{R_r^n}(M_n)$ and $\mathcal{F}^{n+1}$, we have

$$(\mathcal{D}\times\mathcal{F})\backslash(\mathcal{D}^{n+1}\times\mathcal{F}^{n+1})$$
$$=\Big((\mathcal{D}\times\mathcal{F})\backslash(\mathcal{D}^n\times\mathcal{F}^n)\Big)\cup\big(\mathcal{D}^{n+1}\times F_{R_r^n}(M_n)\big)\cup\big(M_n\times\mathcal{F}^{n+1}\big)\cup\big(M_n\times F_{R_r^n}(M_n)\big).$$

So in order to prove (31) at rank $n+1$, we need to show that

$$\mathcal{S}^{n+1} \cap \Big((\mathcal{D} \times \mathcal{F}) \backslash (\mathcal{D}^n \times \mathcal{F}^n) \Big) = \mathcal{S}^n \cap \Big((\mathcal{D} \times \mathcal{F}) \backslash (\mathcal{D}^n \times \mathcal{F}^n) \Big), \tag{32}$$

$$\mathcal{S}^{n+1} \cap \big(M_n \times \mathcal{F}^{n+1} \big) = \mathcal{S} \cap \big(M_n \times \mathcal{F}^{n+1} \big) = \emptyset, \tag{33}$$

$$\mathcal{S}^{n+1} \cap \big(\mathcal{D}^{n+1} \times F_{R_r^n}(M_n) \big) = \mathcal{S} \cap \big(\mathcal{D}^{n+1} \times F_{R_r^n}(M_n) \big) = \emptyset, \tag{34}$$

$$\mathcal{S}^{n+1} \cap \big(M_n \times F_{R_r^n}(M_n) \big) = \mathcal{S} \cap \big(M_n \times F_{R_r^n}(M_n) \big) = \mathcal{S}^0 \cap \big(M_n \times F_{R_r^n}(M_n) \big). \tag{35}$$

To prove these equalities, the main tool is the following formula which is a direct consequence of the construction:

$$\mathcal{S}^{n+1} = \mathcal{S}^n \backslash \big(\mathcal{D}^{n+1} \times F_{R_r^n}(M_n) \big). \tag{36}$$

With this formula at hand, we see that (32) follows from $\mathcal{D}^{n+1} \times F_{R_r^n}(M_n) \subset \mathcal{D}^n \times \mathcal{F}^n$. We also deduce very easily that $\mathcal{S}^{n+1} \cap (\mathcal{D}^{n+1} \times \mathcal{F}^{n+1}) = \mathcal{S}^n \cap (\mathcal{D}^{n+1} \times \mathcal{F}^{n+1}) = \mathcal{S}^0 \cap (\mathcal{D}^{n+1} \times \mathcal{F}^{n+1})$, where the last equality follows from the first point at rank n.

Then, to prove (33), as both $\mathcal{S}$ (by assumption) and $\mathcal{S}^{n+1}$ (by (36)) are included in $\mathcal{S}^n$, it suffices to show that $\mathcal{S}^n \cap (M_n \times \mathcal{F}^{n+1}) = \emptyset$. But that last assertion follows from the definition of $\mathcal{F}^{n+1} = \mathcal{F}^n \backslash F_{R_r^n}(M_n)$: these are precisely the columns where R_r^n has only zero entries on the intersection with the lines M_n.

To prove (34), let us observe that the equality $\mathcal{S}^{n+1} \cap (\mathcal{D}^{n+1} \times F_{R_r^n}(M_n)) = \emptyset$ is a direct consequence of (36). The other equality, namely, $\mathcal{S} \cap (\mathcal{D}^{n+1} \times F_{R_r^n}(M_n)) = \emptyset$ follows from the fact that M_n is a SISP set for $(R_r^n; \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$, so that (29) applies with M_n instead of A, R_r^n instead of R, $\mathcal{D}^n$ instead of $\mathcal{D}$ and $R^{n,*} = R^{*}{}_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}$ instead of R^* (here, we use the point three at rank n). Notice that as $\mathcal{S} \cap (\mathcal{D}^{n+1} \times F_{R_r^n}(M_n)) = \emptyset$ and $\mathcal{S} \subset \mathcal{S}^n$, by (36), we have also proved that $\mathcal{S} \subset \mathcal{S}^{n+1}$.

Finally, to prove (35), as $\mathcal{S} \subset \mathcal{S}^{n+1} \subset \mathcal{S}^0$, we just need to prove that $\mathcal{S} \cap (M_n \times F_{R_r^n}(M_n)) = \mathcal{S}^0 \cap (M_n \times F_{R_r^n}(M_n))$. But this is a direct consequence of the fact that M_n is a SISP set for $(R_r^n; \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$, so that (30) applies with R_r^n instead of R, $R^{n,*}$ instead of R^*, M_n instead of A and $\mathcal{F}^n$ instead of $\mathcal{F}$.

Second point. By definition, $\mathcal{D}^{n+1}$ is empty if and only if $M_n = \mathcal{D}^n$. So if $\mathcal{D}^{n+1} = \emptyset$, then by Assumption 9 applied to $(R_r^n; \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$, we clearly have $F^{R_r^n}(M_n) = \mathcal{F}^n$ and so $\mathcal{F}^{n+1} = \emptyset$. On the other hand, if $\mathcal{D}^{n+1} \neq \emptyset$ we have

$$\begin{aligned} 0 < \mu(\mathcal{D}^{n+1}) &= P^* \Big(\mathcal{D}^{n+1} \times \mathcal{F} \Big) = P^* \Big(\mathcal{D}^{n+1} \times \mathcal{F}^n \Big) \\ &= P^* \Big(\mathcal{D}^{n+1} \times F_{R_r^n}(M_n) \Big) + P^* \Big(\mathcal{D}^{n+1} \times \mathcal{F}^{n+1} \Big) \\ &= P^* \Big(\mathcal{D}^{n+1} \times \mathcal{F}^{n+1} \Big), \end{aligned}$$

where the inequality comes from Assumption 9, the second equality is an easy consequence of (31) at rank n, and the last one comes from the definition of SISP sets (see (29)) and from the fact that P^* and R^* has the same support. So $\mathcal{F}^{n+1}$ cannot be empty, as announced.

Third point. To check that $(R_r^{n+1}; \mu_{\llcorner \mathcal{D}^{n+1}}, \nu_{\llcorner \mathcal{F}^{n+1}})$ satisfies Assumption 9, we need only need to show that the support of $\mu^{R^{n+1}}{}_{\llcorner \mathcal{D}^{n+1}}$ is $\mathcal{D}^{n+1}$ and the support of $\nu^{R^{n+1}}{}_{\llcorner \mathcal{F}^{n+1}}$ is $\mathcal{F}^{n+1}$. But as we already proved that $\mathcal{S} \subset \mathcal{S}^{n+1}$, we know that $P^* \ll R^{n+1}$ and $Q^* \ll R^{n+1}$, so the conclusion follows from the fact that μ and ν have full support by Assumption 9 applied to $(R; \mu, \nu)$. The last thing to check, namely that the matrices $P^{n+1,*}$, $Q^{n+1,*}$ and $R^{n+1,*}$ associated with $(R_r^{n+1}; \mu_{\llcorner \mathcal{D}^{n+1}}, \nu_{\llcorner \mathcal{F}^{n+1}})$ through Theorems 11 and 17 are the restrictions of P^*, Q^* and R^* to $\mathcal{D}^{n+1} \times \mathcal{F}^{n+1}$ is a direct consequence of Proposition 29 below (that we wanted to separate to the rest of the proof because we will use it again later), and of (34). □

Proposition 29. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfy Assumption 8. Let P^*, Q^* and R^* be the matrices associated with the problem $\mathrm{Sch}(R;\mu,\nu)$ by Theorems 11 and 17. Finally, let $A \subset \mathcal{D}$ be such that*

$$R^*\Big((\mathcal{D}\backslash A)\times F_R(A)\Big) = 0. \tag{37}$$

Then (with slightly sloppy notations), ${P^*}_{\llcorner A\times F_R(A)}$, ${Q^*}_{\llcorner A\times F_R(A)}$ and ${R^*}_{\llcorner A\times F_R(A)}$ are the matrices associated with the restricted problem $\mathrm{Sch}(R_{\llcorner A\times F_R(A)}, \mu_{\llcorner A}, \nu_{\llcorner F_R(A)})$ by Theorem 11 and 17.

Similarly, calling $A' := \mathcal{D}\backslash A$ and $F' := \mathcal{F}\backslash F_R(A)$, ${P^*}_{\llcorner A'\times F'}$, ${Q^*}_{\llcorner A'\times F'}$ and ${R^*}_{\llcorner A'\times F'}$ are the matrices associated with the restricted problem $\mathrm{Sch}(R_{\llcorner A'\times F'}, \mu_{\llcorner A'}, \nu_{\llcorner F'})$ by Theorem 11 and 17.

Proof. We show the result in the case of P^*, related to the "block" $A\times F_R(A)$. The case of Q^* is similar, the case of R^* easily follows from the two previous ones, and the similar results on $A'\times F'$ follow the same lines. Let ν^* be defined by (13). The first thing to prove is

$$\begin{aligned}&{\nu^*}_{\llcorner F_R(A)}\\ &= \arg\min\Big\{H(\bar\nu|\nu_{\llcorner F_R(A)})\,\Big|\,\bar\nu = \nu^P \text{ for some } P \text{ with } H(P|R_{A\times F_R(A)}) < +\infty \text{ and } \mu^P = \mu_{\llcorner A}\Big\}.\end{aligned}$$

The measure ${\nu^*}_{\llcorner F_R(A)}$ is a competitor for the problem in the r.h.s. because it corresponds to $P := {P^*}_{\llcorner A\times F_R(A)}$. Let us show that it is the optimizer. To do this, we call $\bar\nu$ the optimizer, and we show that $\nu^*_{F_R(A)} = \bar\nu^*$. Let us consider $\bar P$ a P corresponding to $\bar\nu^*$ in the problem above, $\bar P^*$ the matrix obtained by replacing the entries of P^* on $A\times F_R(A)$ by the entries of $\bar P$, and $\bar\nu^* := \nu^{\bar P^*}$. We have

$$H(\bar\nu^*|\nu) = H(\bar\nu|\nu_{\llcorner F_R(A)}) + H({\nu^*}_{\llcorner F'}|\nu_{\llcorner F'}) \le H({\nu^*}_{\llcorner F_R(A)}|\nu_{\llcorner F_R(A)}) + H({\nu^*}_{\llcorner F'}|\nu_{\llcorner F'}) = H(\nu^*|\nu),$$

where the inequality, being a consequence of the optimality of $\bar\nu^*$, is an equality if and only if $\nu^*_{F_R(A)} = \bar\nu^*$. But by optimality of ν^* in (13) this inequality is indeed an equality, and therefore $\nu^*_{F_R(A)} = \bar\nu^*$.

It remains to show that ${P^*}_{\llcorner A\times F_R(A)}$ is the solution of $\mathrm{Sch}(R_{\llcorner A\times F_R(A)}, \mu_{\llcorner A}, {\nu^*}_{\llcorner F_R(A)})$. For this, let us consider $\bar P$ the solution of $\mathrm{Sch}(R_{\llcorner A\times F_R(A)}, \mu_{\llcorner A}, {\nu^*}_{\llcorner F_R(A)})$, and $\bar P^*$ the matrix obtained by replacing the entries of P^* on $A\times F_R(A)$ by the entries of $\bar P$. Because of (37), we have

$$\begin{aligned}H(\bar P^*|R) &= H(\bar P|R_{\llcorner A\times F_R(A)}) + H({P^*}_{\llcorner A'\times F'}|R_{\llcorner A'\times F'})\\ &\le H({P^*}_{\llcorner A\times F_R(A)}|R_{\llcorner A\times F_R(A)}) + H({P^*}_{\llcorner A'\times F'}|R_{\llcorner A'\times F'}) = H(P^*|R^*),\end{aligned}$$

where the inequality is a consequence of the optimality of $\bar P$, and is an equality if and only if $\bar P = {P^*}_{\llcorner A\times F_R(A)}$. But by optimality of P^*, this inequality is indeed an equality, and we conclude that $\bar P = {P^*}_{\llcorner A\times F_R(A)}$. The proposition is proved. □

Now, we want to prove Lemma 27. To do this, we introduce a new class of subsets of $\mathcal{D}$, associated with a triple $(R;\mu,\nu)$.

Definition 30. Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9. The maximal θ associated to $(R;\mu,\nu)$ is defined by:

$$\theta_m := \max_{\substack{A\subset\mathcal{D}\\ A\neq\emptyset}} \frac{\mu(A)}{\nu(F_R(A))}. \tag{38}$$

We say that $A \subset \mathcal{D}$ is a maximal θ-set for $(R;\mu,\nu)$ if A is a maximizer of (38). We say that it is a *smallest* maximal θ-set if in addition, it is a minimal element in the sense of inclusion among all maximal θ-sets associated with $(R;\mu,\nu)$.

As maximal θ-sets are optimizers of a finite function (thanks to Assumption 9) on a finite set (the set of all nonempty subsets of $\mathcal{D}$), any triple $(R;\mu,\nu)$ satisfying Assumption 9 admits at least one maximal θ-set. The set of all maximal θ-sets being itself finite, we know that there exists at least one minimal element in this set, so that smallest maximal θ-sets always exist under Assumption 9. Hence, Lemma 27 is an obvious consequence of the following proposition, whose proof heavily relies on the optimality conditions stated in Proposition 19.

Proposition 31. *Let $R \in \mathcal{M}_+(\mathcal{D}\times\mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9. A smallest maximal θ-set for $(R;\mu,\nu)$ is a SISP set for $(R;\mu,\nu)$.*

Proof. Let μ^* and ν^* be defined by (13). We first define the two following quantities

$$\theta^{\mathcal{D}} := \max_{i\in\mathcal{D}} \frac{\mu_i}{\mu_i^*}, \qquad \theta^{\mathcal{F}} := \max_{j\in\mathcal{F}} \frac{\nu_j^*}{\nu_j}.$$

(Recall that under Assumption 9, by Remark 13, μ^* and ν^* defined by (13) have full support.) Then, we define the two following sets, that are nonempty subsets of $\mathcal{D}$ and $\mathcal{F}$ respectively:

$$\overline{M} := \left\{x_i \in \mathcal{D} \text{ s.t. } \frac{\mu_i}{\mu_i^*} = \theta^{\mathcal{D}}\right\}, \qquad \overline{F} := \left\{y_j \in \mathcal{F} \text{ s.t. } \frac{\nu_j^*}{\nu_j} = \theta^{\mathcal{F}}\right\}.$$

The main argument of the proof consists in showing that $\theta^{\mathcal{D}}$ and $\theta^{\mathcal{F}}$ coincide with θ_m, the maximal θ for $(R;\mu,\nu)$. Even if $\overline{M}$ is not a smallest maximal-θ set in general (more precisely, it is a maximal θ-set that is not minimal in general), we show at Step 3 below how this information allows to conclude. As before, P^* is the matrix defined by (14).

Step 1: $\theta^{\mathcal{D}} = \theta^{\mathcal{F}}$.

Let $x_i \in \overline{M}$ and $y_j \in \mathcal{F}$ be such that $P^*_{ij} > 0$ (such a j exists thanks to Assumption 9). By the first line of (26), we have

$$\frac{\mu_i^*}{\mu_i}\frac{\nu_j^*}{\nu_j} = 1, \tag{39}$$

which implies that $\nu_j^*/\nu_j = \theta^{\mathcal{D}}$, and hence that $\theta^{\mathcal{F}} \geq \theta^{\mathcal{D}}$. The other inequality is proved in the same way, and the result follows. From now on, we call

$$\overline{\theta} := \theta^{\mathcal{D}} = \theta^{\mathcal{F}}.$$

Step 2: $\overline{\theta} = \theta_m$.

First, $\overline{\theta} \geq \theta_m$. Indeed, for any $A \subset \mathcal{D}$, by Theorem (23), as $\mathrm{Sch}(R;\mu,\nu^*)$ is at least approximately scalable, we have $\mu(A) \leq \nu^*(F_R(A))$. But on the other hand, by definition of $\overline{\theta}$, we have $\nu^* \leq \overline{\theta}\nu$, so that actually, $\mu(A) \leq \overline{\theta}\nu(F_R(A))$, and hence $\overline{\theta} \geq \theta_m$.

Also, $\overline{\theta} \leq \theta_m$. To see this, let us first observe that $P^*((\mathcal{D}\backslash\overline{M})\times\overline{F}) = P^*(\overline{M}\times(\mathcal{F}\backslash\overline{F})) = 0$. This is because if x_i, y_j are such that $P^*_{ij} > 0$, still by (39), $\mu_i/\mu_i^* = \overline{\theta}$ if and only if $\nu_j^*/\nu_j = \overline{\theta}$, so that $x_i \in \overline{M}$ if and only if $y_j \in \overline{F}$. Therefore, on the one hand, $F_R(\overline{M}) \subset \overline{F}$, and on the other hand, projecting on both marginals:

$$\begin{aligned}\mu(\overline{M}) &= P^*(\overline{M}\times\mathcal{F}) = P^*(\overline{M}\times\overline{F}) + \underbrace{P^*(\overline{M}\times(\mathcal{F}\backslash\overline{F}))}_{=0} \\ &= P^*(\overline{M}\times\overline{F}) + \underbrace{P^*((\mathcal{D}\backslash\overline{M})\times\overline{F})}_{=0} = P^*(\mathcal{D}\times\overline{F}) = \nu^*(\overline{F}).\end{aligned}$$

As by definition of $\overline{F}$ and $\overline{\theta}$, $\nu^*(\overline{F}) = \overline{\theta}\nu(\overline{F})$, we conclude that

$$\overline{\theta}\nu(F_R(\overline{M})) \leq \overline{\theta}\nu(\overline{F}) = \mu(\overline{M}),$$

so that $\overline{\theta} \leq \theta_m$, as announced.

Step 3: Conclusion.

We are now in position to conclude. Let A be a smallest maximal θ-set. As $\mathrm{Sch}(R;\mu,\nu^*)$ is at least approximately scalable, we know that $\mu(A) \geq \nu^*(F_R(A))$. On the other hand, as A is a maximal θ-set, we know that $\mu(A) = \theta_m\nu(F_R(A)) = \overline{\theta}\nu(F_R(A))$. But by definition of $\overline{\theta}$, we know that $\overline{\theta}\nu \geq \nu^*$, so that $\nu^*(F_R(A)) \leq \mu(A)$. We conclude that $\nu^*(F_R(A)) = \mu(A)$, that ${\nu^*}_{\llcorner A} = \overline{\theta}\nu_{\llcorner A} = \theta_m\nu_{\llcorner A}$, and hence that

$$P^*\Big((\mathcal{D}\backslash A)\times F_R(A)\Big) = P^*(\mathcal{D}\times F_R(A)) - P^*(A\times F_R(A)) = \nu^*(F_R(A)) - \mu(A) = 0,$$

so that (29) holds.

In addition, by Proposition 29 the measure ${P^*}_{\llcorner A\times F_R(A)}$ is the solution of the problem $\mathrm{Sch}(R_{\llcorner A\times F_R(A)};\mu_{\llcorner A},{\nu^*}_{\llcorner F_R(A)})$. So in order to prove (35), it suffices to prove that this problem is scalable. For this purpose, we will use Lemma 24. We call $R_r := R_{\llcorner A\times F_R(A)}$. Let B be a nonempty strict subset of A. As A is a minimal element in the set of maximal θ-sets for $(R;\mu,\nu)$, we know that $\mu(B) < \theta_m\nu(F_R(B))$. Then, as $F_R(B) \subset F_R(A)$, we have $F_R(B) = F_{R_r}(B)$, so $\mu(B) < \theta_m\nu(F_{R_r}(B))$. Finally, as ${\nu^*}_{\llcorner A} = \theta_m\nu_{\llcorner A}$, we have $\mu(B) < \nu^*(F_{R_r}(B))$. So Lemma 24 applies, and $\mathrm{Sch}(R_{\llcorner A\times F_R(A)};\mu_{\llcorner A},{\nu^*}_{\llcorner F_R(A)})$ is scalable, which concludes the proof. □

We close this section with a remark concerning the stability with respect to union of SISP sets.

Remark 32. It is easy to check that SISP sets associated with a triple $(R;\mu,\nu)$ are stable by union. Therefore, there exists an upper bound in the set of all SISP sets for $(R;\mu,\nu)$, that we call the *largest* SISP set. If we want the procedure described in Proposition 28 to be as fast as possible, it is logical to look for SISP sets that are as large as possible, in order to minimize the rank N at which the procedure reaches its stationary point. This is what we are going to do in the next section.

6 Numerical applications

A simple consequence of the theoretical procedure described in the previous section is that in a lot of cases, if the problem $\mathrm{Sch}(R;\mu,\nu)$ is non-scalable, the matrices P^* and Q^* from Theorem 11, and R^* from Theorem 17 have more zero entries than R. For instance, if the problem is balanced (*i.e.* $\mathsf{M}(\mu) = \mathsf{M}(\nu)$), and if the bipartite graph of R is connected (that is, $\mu(A) = \nu(F_R(A))$ only holds for $A = \emptyset$ or $A = \mathcal{D}$, which is a reasonable assumption in a lot of contexts), we can check that the matrix R^1 from Proposition 28 cannot coincide with R.

Therefore, typically, the Sinkhorn algorithm in the non-scalable case does not converge linearly. We are going to detail an approximate algorithm allowing to find the common support of P^*, Q^* and R^*, and therefore to recover a linear rate of convergence for the Sinkhorn algorithm by Proposition 25.

6.1 Stopping criterion

Before any numerical application, we need to define a stopping criterion for the Sinkhorn algorithm when the Schrödinger problem is non-scalable. When the problem is scalable, the classical criterion that is used is the duality gap estimated at each step $n \in \mathbb{N}^*$ of the Sinkhorn algorithm:

$$SC^n = H(P^n|R) - \langle \log(a^n),\mu\rangle - \langle \log(b^{n-1}),\nu\rangle, \tag{40}$$

where P^n, a^n and b^{n-1} are defined by the relations (2). Indeed, it is known that this quantity is always positive when the relation $P^n_{ij} = a^n_i b^{n-1}_j R_{ij}$ holds for all i,j, and the Fenchel-Rockafellar

duality ensures that $SC_n \to 0$ as $n \to \infty$ when the problem is scalable, *i.e.* when (a^n) and (b^n) converge.

In the approximately scalable case, numerical instabilities may appear when $n \to \infty$ because (a^n) and (b^n) do not converge, but this criterion may remain useful if the error that is tolerated is not too small. However, in the non-scalable case, this criterion does not hold as the problem $\mathrm{Sch}(R;\mu,\nu)$ has no solution. The results presented in the previous sections allow nevertheless to define an approximate criterion. Indeed, it has been shown in [6] that for a given $\lambda > 0$, the problem defined with the notations of Subsection 2.1 by

$$\mathrm{Schu}_\lambda(R;\mu,\nu) = \min\left\{H(P|R) + \lambda H(\nu^P|\nu) \,\middle|\, P \text{ s.t. } \mu^P = \mu\right\}, \tag{41}$$

can be solved numerically with a generalization of the Sinkhorn algorithm. More precisely, the duality gap defined for all $n \in \mathbb{N}^*$ by

$$SCu^n_\lambda = H(P^n|R) + \lambda\left(H(\nu^{P^n}|\nu) - \left\langle 1 - (1/b^{n-1})^{\frac{1}{\lambda}}, \nu\right\rangle\right) - \langle \log(a^n), \mu\rangle, \tag{42}$$

converges to 0 whenever $P^n_{ij} := a^n_i b^{n-1}_j R_{ij}$ for all i,j, and with a^n, b^n defined for all $n \in \mathbb{N}^*$ by the relations:

$$\begin{cases} \forall j, \quad b^0_j := 1, \\ \forall n \geq 0, \quad \forall i, \quad a^{n+1}_i := \dfrac{\mu_i}{\sum_j b^n_j R_{ij}}, \\ \forall n \geq 0, \quad \forall j, \quad b^{n+1}_j := \left(\dfrac{\nu_j}{\sum_i a^{n+1}_i R_{ij}}\right)^{\frac{\lambda}{1+\lambda}}. \end{cases} \tag{43}$$

On the other hand, a slight modification of our Γ-convergence result of Proposition 15 asserts that P^*, from Theorem 11, is the limit of the solution of (41) as $\lambda \to +\infty$. So if we now define SCu^n_λ by the formula (42) where P^n, a^n and b^{n-1} are computed with the standard Sinkhorn algorithm (2), instead of the modified one (43), we conclude that for all $\varepsilon > 0$, there exists a threshold λ^ε such that for all $\lambda \geq \lambda^\varepsilon$,

$$\limsup_{n \to +\infty} SCu^n_\lambda \leq \varepsilon.$$

Therefore, the stopping criterion (42) can still be used for the sequence $(P^n)_{n\in\mathbb{N}}$ generated by the classical Sinkhorn algorithm (2), as long as λ^ε is chosen to be sufficiently large w.r.t. ε. In practice, we observe that taking $\lambda^\varepsilon = \frac{1}{\varepsilon}$ works well. This is what we are going to do in the following section, considering a level of error $\varepsilon := 10^{-3}$.

6.2 An approximate numerical method for constructing the support of R^*

An interesting application of the theoretical procedures described in Subsection 5.2 is the construction of an approximate algorithm allowing the identification of the support $\mathcal{S}$ of R^* w.r.t. R, when the problem $\mathrm{Sch}(R;\mu,\nu)$ is approximately scalable or non-scalable. Our motivation is double. First, the fact that knowing the support allows to recover a linear rate for the Sinkhorn algorithm, thanks to Proposition 25, suggests that if this algorithm is simple enough, the full procedure to obtain R^* (preprocessing to know the $\mathcal{S}$ and then the Sinkhorn algorithm applied to $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}}R;\mu,\nu)$ is excepted to be faster than the Sinkhorn algorithm applied directly to $\mathrm{Sch}(R;\mu,\nu)$. Second, even if we have shown that the Sinkhorn algorithm converges in the non-scalable case, the Schrödinger potentials appearing in the Sinkhorn procedure are likely to be too high to be computed numerically before the algorithm has converged: thus,

finding the support before to run the Sinkhorn algorithm can be an advantage, as long as this preprocessing prevents this potential's explosion and even if it does not accelerate the procedure.

We recall that we identified, at each iteration n of the procedure described in Proposition 28, a SISP set, denoted M_n, for a reduced problem $\mathrm{Sch}(R_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}; \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$. Regarding the proof of Lemma 27, a natural choice for M_n is an union of smallest maximal θ-sets for $\mathrm{Sch}(R_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}; \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$, that we introduced in Definition 30. The approximate procedure that we design for finding M_n at each iteration consists in two steps:

1. Find the largest set M'_n (in the sens of inclusion) such that the supports of R and R^* coincide in $M'_n \times \mathcal{F}$;
2. Find $M_n \subset M'_n$, an union of smallest maximal θ-set for $(R_{\llcorner M'_n \times F_R(M'_n)}; \mu_{\llcorner M'_n}, \nu_{\llcorner F_R(M'_n)})$.

Indeed, finding M'_n in the first step can be achieved with a slightly modified Sinkhorn procedure applied to
$\mathrm{Sch}(R_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}, \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$. More precisely, we initialize $U = \mathcal{D}^n$ and $V = \mathcal{F}^n$, and at each step of this Sinkhorn procedure, for all $x_i \in \mathcal{D}^n$, if there exists $y_j \in V$ such that the coupling obtained at this step is smaller than a given threshold, we set $U = U \backslash x_i$. Since Lemma 27 ensures that there exists at least one SISP set, M'_n is not empty and thus we do not set all the lines to 0 (given that the threshold is sufficiently small). For all $y_j \in V$, if $y_j \notin F_{R_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}}(U)$, we set $V = V \backslash y_j$.

After several steps of this modified Sinkhorn procedure, the elements remaining in U and V coincide respectively with M'_n and $F_{R_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}}(M'_n)$. Since the Sinkhorn algorithm restricted to $U \times V$ converges linearly (thanks to Proposition 25), we then rapidly observe the convergence of the procedure.

For detailing how to obtain M_n at the second step, we need to introduce the notion of connected components:

Definition 33. Let $U \subset \mathcal{D}$ and $V \subset \mathcal{F}$. We say that $A \times B$ is a connected component of the graph $(U \cup V, \triangle)$ (using the notation of 28) if:

- $R_{\llcorner U \times V}(A \times B^c) = R_{\llcorner U \times V}(A^c \times B) = 0$;
- $(A \cup B, \triangle)$ is connected.

Finding connected components for such undirected graph is a classical task in Graph theory, for which there exists ready-to-use algorithms [12]. We show in the following proposition that we can define at the second step of the procedure

$$M_n = \bigcup \{U_i, \mid \frac{\mu(U_i)}{\nu(V_i)} = \max_{j=1,\cdots,C} \frac{\mu(U_j)}{\nu(V_j)}, \tag{44}$$

where $(U_1 \times V_1, \cdots, U_C \times V_C)$ are the C connected components of the graph $(M'_n \cup F_{R_{\llcorner \mathcal{D}^n \times \mathcal{F}^n}}(M'_n), \triangle)$.

Proposition 34. *Let $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ satisfying Assumption 9, and let us denote U the largest subset of $\mathcal{D}$ (in the sens of inclusion) such that the supports of R and R^* coincide in $U \times \mathcal{F}$, and $V = F_R(U)$. Let us denote $(U_1 \times V_1, \cdots, U_C \times V_C)$ the C connected components of the graph $(U \cup V, \triangle)$. Then, for any $U_i \times V_i$ such that*

$$\frac{\mu(U_i)}{\nu(V_i)} = \max_{j=1,\cdots,C} \frac{\mu(U_j)}{\nu(V_j)}, \tag{45}$$

U_i is a smallest maximal θ-sets for $(R; \mu, \nu)$.

Proof. The main point consists in showing that a smallest maximal θ-set for $(R;\mu,\nu)$, that we denote A is such that $A \times F_R(A)$ is a connected component of $(U \cup V, \triangle)$. Indeed, assuming this claim, the fact that there always exists such smallest maximal θ-set ensures that for any $U_i \times V_i$ maximizing (45), U_i is a maximal θ-set. U_i is thus necessarily a smallest one because it cannot contain any other connected component, and then any smallest maximal θ-set neither. We now prove the claim.

As A is a SISP set (thanks to the proof of Proposition 27), it is in U and we have necessarily $R^*(A^c \times F_R(A)) = R^*(A \times F_R(A)^c) = 0$. The supports of R and R^* being the same in $U \times V$ by construction of U and V, we have thus $R_{\llcorner U\times V}(A^c \times F_R(A)) = R_{\llcorner U\times V}(A \times F_R(A)^c) = 0$. Moreover, $A \times F_R(A)$ cannot contain any connected component of $(U \cup V, \triangle)$: if it was the case, this connected component would characterize a maximal θ-set for $(R;\mu,\nu)$, which would contradict the minimality of A. Thus, $A \times F_R(A)$ is necessarily a connected component of the graph $(U \cup V, \triangle)$. □

As we already showed that Assumption 9 holds for $(R_{\llcorner \mathcal{D}^n\times\mathcal{F}^n}, \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$ for every n, we can apply Proposition 34 to this triple. Moreover, as we have seen in the proof that any smallest maximal θ-set characterizes a connected component of $(U \cup V, \triangle)$, the choice of M_n defined by (44) corresponds in fact to the union of all the smallest maximal θ-sets for $(R_{\llcorner \mathcal{D}^n\times\mathcal{F}^n}, \mu_{\llcorner \mathcal{D}^n}, \nu_{\llcorner \mathcal{F}^n})$.

We provide in Algorithm 1 the pseudo-code of this iterative method.

Let us make a few comment on this algorithm. The support of R^* only depends on the support of R and not of its values, so when identifying $\mathcal{S}$, we can equivalently consider the problem $\mathrm{Sch}(R';\mu,\nu)$, where $R' = \mathbb{1}_{R\neq 0}$. This explains why we consider the minimum $\min_{y_j\in V} a_i \times b_j$ rather than $\min_{y_j\in V} a_i \times b_j \times R_{ij}$ in Algorithm 1.

The stopping criterion corresponds to the criterion (42) detailed in the previous section, which has to be smaller than a certain threshold ε to be satisfied.

Choosing in an appropriate way the set of minimal factors $\{m_i,\, i = 1, \cdots, N\}$, is crucial: it determines the level of approximation that is considered as acceptable, *i.e.* the minimal value at which we can consider that the algorithm should create a new zero entries. In practice, we observe that

$$m_i := \frac{1}{n}\frac{\mu_i}{\mu_i^R}, \tag{46}$$

seems to be a good tradeoff between efficiency and security in most of the cases that we explored.

Remark 35.

- This method can be seen as an improvement over the naive approximate method which consists, at each iteration of the Sinkhorn algorithm applied to $\mathrm{Sch}(R;\mu,\nu)$, to set all the entries of R that are smaller to a certain threshold to zero. With our method, we do not have to identify all the zero entries one by one but line by line, which can avoid numerical errors which may appear for some entries converging slowly to zero. However, this is done at the cost of identifying the connected components of some subgraphs of $(\mathcal{D} \cup \mathcal{F}, \triangle)$ at every iteration of the procedure, which slows down the algorithm in cases where these subgraphs are large. Note that in typical cases where the matrix R is structured, we do not expect to find more than one connected component at each iteration as in Figure 4.

- We emphasize the fact that this algorithm is only approximate, and this for two reasons. The first one occurs when the set of thresholds $\{m_i,\, i = 1, \ldots, N\}$ are too large. Then, we can set to zero lines which should not be, just because some of the entries of R^*

Algorithm 1 Find the support Supp of R^*: return Supp

Require: • A set of minimal factors: $\{m_i,\, i = 1, \cdots, N\}$,
• A stopping criterion: $stop(a, b, R, \mu, \nu)$.
We set $A = \mathcal{D}$, $B = \mathcal{F}$, Supp = Support(R).
while $A \neq \emptyset$ **do**
$\bar{R} = R_{\llcorner A\times B}$, $\bar{\mu} = \mu_{\llcorner A}$, $\bar{\nu} = \nu_{\llcorner B}$
$b = \mathbb{1}_B$, $a = \mathbb{1}_A$
$U = A$, $V = B$
while $stop(a, b, \bar{R}, \bar{\mu}, \bar{\nu}) \neq 1$ **do**
for $x_i \in U$ **do**
if $\sum_{y_j \in V} \bar{R}_{ij} b_j = 0$ **then**
$U = U \backslash \{x_i\}$
else
$a_i = \dfrac{\bar{\mu}_i}{\sum_{y_j \in V} \bar{R}_{ij} b_j}$
if $\min_{y_j \in V} a_i \times b_j < m_i$ **then**
$U = U \backslash \{x_i\}$
end if
end if
end for
$\bar{R} = \bar{R}_{\llcorner U\times V}$, $\bar{\mu} = \bar{\mu}_{\llcorner U}$, $a = a_{\llcorner U}$
for $y_j \in V$ **do**
if $\sum_{x_i \in A} \bar{R}_{ij} a_i = 0$ **then**
$V = V \backslash \{y_j\}$
else
$b_j = \dfrac{\bar{\nu}_j}{\sum_{x_i \in A} \bar{R}_{ij} a_i}$
end if
end for
$\bar{R} = \bar{R}_{\llcorner U\times V}$, $\bar{\nu} = \bar{\nu}_{\llcorner V}$, $b = b_{\llcorner V}$
end while
$(U_1 \times V_1, \cdots, U_C \times V_C)$ = connected components of the graph $(U \cup V, \triangle)$
$U \times V = \bigcup \big\{ U_i \times V_i, \mid \frac{\mu(U_i)}{\nu(V_i)} = \max_{j=1,\cdots,C} \frac{\mu(U_j)}{\nu(V_j)} \big\}$
$A = A \backslash U$, $B = B \backslash V$
Supp = Supp$\backslash \big((A \backslash U) \times V \big)$
end while

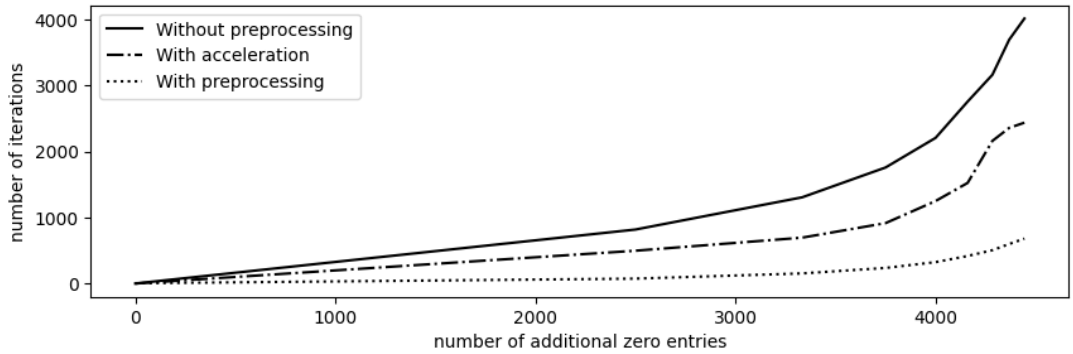


Figure 5 : Number of iterations needed for convergence *vs.* number of additional zero entries of the limits P^* and Q^* from Theorem 11 w.r.t. R, using: the Sinkhorn algorithm (2) (solid line); the Sinkhorn algorithm where at each step, the entries below the threshold (46) are set to zero (dashed line); Algorithm 1 to compute the support $\mathcal{S}$ of the limits, and then the Sinkhorn algorithm replacing R by $\mathbb{1}_{\mathcal{S}}R$ (dotted line). When the threshold ε is small enough, as it is the case here, these three methods provide the same limits.

should be small on this line. This must be avoided as then the algorithm cannot converge towards R^*.

The other case where our algorithm does not identify $\mathcal{S}$ exactly is either when the threshold ε of the stopping criterion is large, or when the thresholds $\{m_i,\, i = 1, \dots, N\}$ are small. Then, the Sinkhorn algorithm can satisfy the stopping criterion before all the zeroes have been identified. This is not a big problem, since it means that the algorithm converges well without having to identify the additional zeroes of R^*.

With these observations, we conclude that the thresholds $\{m_i,\, i = 1, \dots, N\}$ need to be taken rather small w.r.t. the threshold ε of the stopping criterion, even though of course, if they are taken too small, efficiency is lost since then the algorithm just behaves as Sinkhorn without any improvement.

We illustrate in Figure 5 the efficiency of this procedure. As emphasized at the beginning of this section, the number of iterations of a Sinkhorn-like algorithm is a better indicator than the computation time of the full procedure because our main motivation to find the support of R^* before running the Sinkhorn algorithm is that we want the Schrödinger potentials not to explode before reaching convergence, which could be the case for non-scalable problems. We thus represent the number of iterations needed for the Sinkhorn algorithm to converge as a function of the number of additional zero entries in R^* w.r.t. to R, and compare when we apply or not the preprocessing described in Algorithm 1. For varying the number of zero entries, we take R upper-diagonal, build μ and ν similarly to what we described in Figure 4, and then vary the number of blocks from 1 (corresponding to the scalable case) to 10. For the case with preprocessing, we consider the sum of the iterations needed for the Sinkhorn-like method described in Algorithm 1 to find the support $\mathcal{S}$, and of the ones needed for the Sinkhorn algorithm then applied to the problem $\mathrm{Sch}(\mathbb{1}_{\mathcal{S}}R; \mu, \nu)$. We observe that the preprocessing makes the number of iterations needed for the convergence to be significantly smaller than for the case without preprocessing when the number of additional zero entries is high. It is also smaller than for the case when the naive approximate method is applied, illustrating the benefit of our approach.

6.3 Comparison of the method with the balanced and unbalanced Sinkhorn algorithms

We compared in Figure 6 the outputs of the Sinkhorn algorithm when the problem $\mathrm{Sch}(R;\mu,\nu)$ is non-scalable, given by the geometric mean described in Theorem 17, and two alternatives:

- When the reference coupling R is modified such that it has only positive entries. For that, we built a new coupling R_ε by adding on every zero entry of R a small quantity ε. We then found the optimizers R^*_ε of the Schrödinger problems $\mathrm{Sch}(R_\varepsilon;\mu,\nu)$ and compared its distance in total variation of its solution to R^*, for different values of ε.
- When we the marginal constraints are replaced by marginal penalizations, leading to an unbalanced problem of the form (21), using the scaling algorithm described in [6]. We then compared the distance in total variation of its solution to R^*, for different values of λ.

For the comparison realized here, we took a coupling R of size 100×100 and R,μ,ν as in Figure 4 in such a way that R^* has only two blocks $A_1\times B_1$ and $A_2\times B_2$ for which the factor λ appearing in the procedure of Proposition 28 is greater than one for the first component, and smaller for the second one. The problem is thus non-scalable. As expected, we observe in Figure 6a that in the first case it is impossible for the solution R^*_ε of $\mathrm{Sch}(R_\varepsilon;\mu,\nu)$ to be close to R^* (and then to recover the right minimum entropy), and that the faster is the convergence, the further R^*_ε is from R^*. In the second case, we observe in Figure 6b that the solution R^*_λ of the unbalanced problem with penalization λ converges to R^* when $\lambda\to\infty$. However, the convergence goes faster only for values of λ smaller than 150, for which we still observe a significant difference between R^* and R^*_λ.

Note that these results are not limited to Schrödinger problems similar to the one described in Figure 4, and that we observed the same type of results for randomly-generated R, μ and ν in non-scalable cases.

Acknowledgments

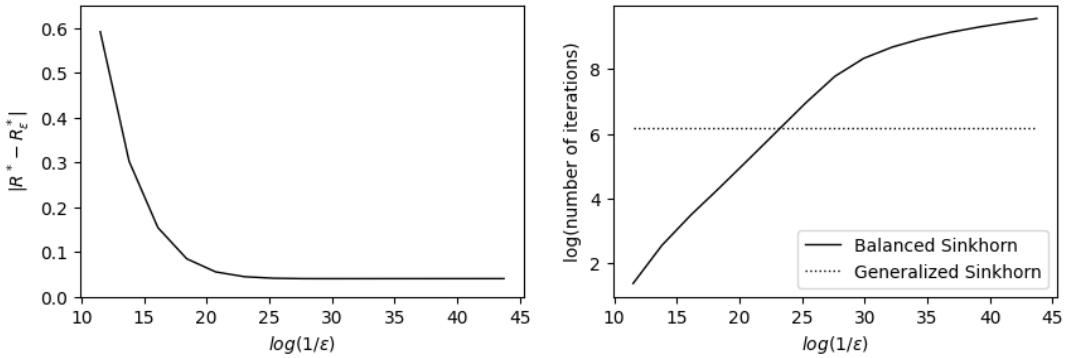
This work was supported by funding from French agency ANR (SingleStatOmics; ANR-18-CE45-0023-03). We would like to thank Olivier Gandrillon, Thibault Espinasse and Thomas Lepoutre for enlightening discussions, and the latter and Gauthier Clerc for critical reading of the manuscript. We finally thank the BioSyL Federation, the LabEx Ecofect (ANR-11-LABX-0048) and the LabEx Milyon of the University of Lyon for inspiring scientific events.

Appendices

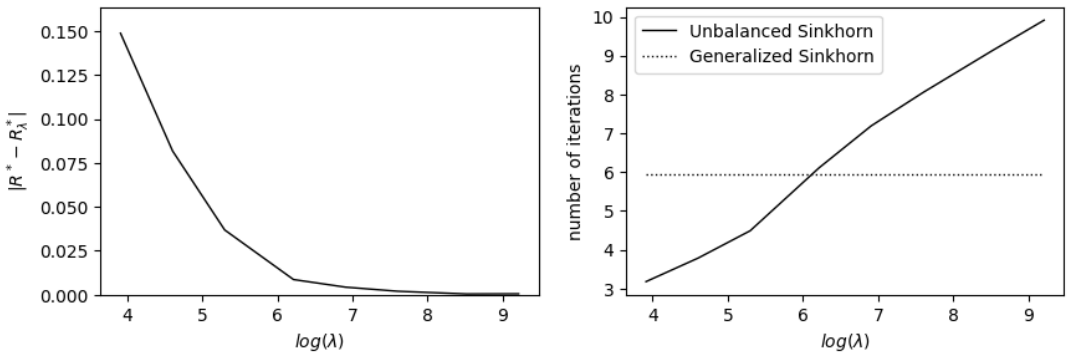
A Example of Schrödinger problems without solutions

There exists a lot of degenerate cases where the problem $\mathrm{Sch}(R;\mu,\nu)$ has no solution. Indeed, in the extreme situation where most of the entries of R cancel, two randomly chosen vectors μ and ν have more chance to be non-scalable than to satisfy the conditions of Theorem 23. For example, in the typical example of a squared diagonal reference coupling R, we must necessarily have $\mu=\nu$ for these conditions to be satisfied.

When illustrating our results at Sections 5 and 6, we chose R to be a squared upper-diagonal matrix (see Figure 4). This is of particular interest, as it corresponds to a case that typically arises when considering entropy minimization problems in cell biology. Indeed, the dynamics of mRNA levels within a cell, which drives cellular differentiation processes, is often modeled by a piecewise deterministic Markov process, where stochastic bursts of mRNAs compensate



(a)



(b)

Figure 6 : Comparison of the outputs of the Sinkhorn algorithm in a non-scalable case, where R^* is given by Theorem 17, and: 6a the Sinkhorn algorithm (2) where the zero entries of R are replaced by a small value ε; 6b the unbalanced Sinkhorn algorithm from [6] applied to solve $\mathrm{Schu}_\lambda(R; \mu, \nu)$ for large values of λ.

their deterministic degradation [32, 33]. Considering the simplest cartoonish but enlightening situation where there is no degradation, a constant number of cells, and where we measure the activity of only one gene, the quantity of mRNAs in the cells corresponding to this gene can only increase with time. Therefore, if R is the matrix whose entry R_{ij} gives the number of cells having i molecules of mRNA at a first timepoint and j molecules at a later timepoint, R must be upper-diagonal.

To give an insight of the behaviour of the Sinkhorn algorithm in the non-scalable case with an upper-diagonal reference matrix, let us treat explicitly a simple example. We consider:

$$R = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mu = (2,2,2), \quad \nu = (2,3,1).$$

In this example, $\mu_3 > \nu_3$ while the image of x_3 by the graph associated to R is reduced to y_3, that is, with the formalism of Section 5, $F_R(\{x_3\}) = \{y_3\}$ and hence $\nu(F_R(\{x_3\}) < \mu(\{x_3\})$. In view of Theorem 23 (which is very easy to check in our simple situation), the problem is therefore indeed non-scalable: no matrix can satisfy the marginal constraints and be absolutely continuous w.r.t R at the same time.

With the notations of (2), let us reproduce below the output of the Sinkhorn algorithm at some of the first iterations. Starting at Iteration 5, we only give approximate numerical values.
Iteration 1:

$$a^1 = (2/3,1,2), \quad b^0 = (1,1,1), \quad P^1 = \begin{pmatrix} 2/3 & 2/3 & 2/3 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Iteration 2:

$$a^1 = (2/3,1,2), \quad b^1 = (3,9/5,3/11), \quad Q^1 = \begin{pmatrix} 2 & 6/5 & 2/11 \\ 0 & 9/5 & 3/11 \\ 0 & 0 & 6/11 \end{pmatrix}.$$

Iteration 5:

$$a^3 = (2.7e^{-1}, 8.6e^{-1}, 1.7e^{1}), \quad b^2 = (5.0, 2.2, 1.2e^{-1}), \quad P^3 = \begin{pmatrix} 1.4 & 0.59 & 3.2e^{-2} \\ 0 & 1.9 & 1.0e^{-1} \\ 0 & 0 & 2.0 \end{pmatrix}.$$

Iteration 11:

$$a^6 = (1.2e^{-1}, 5.1e^{-1}, 1.5e^{2}), \quad b^5 = (1.3e^{1}, 3.9, 1.3e^{-2}), \quad P^6 = \begin{pmatrix} 1.55 & 0.45 & 1.5e^{-3} \\ 0 & 2.0 & 6.6e^{-3} \\ 0 & 0 & 2.0 \end{pmatrix}.$$

Iteration 80:

$$a^{40} = (5.5e^{-5}, 2.8e^{-4}, 2.7e^{12}), \quad b^{40} = (3.6e^{4}, 9.1e^{3}, 3.8e^{-13}), \quad Q^{40} = \begin{pmatrix} 2.0 & 0.50 & 1.7e^{-17} \\ 0 & 2.5 & 8.4e^{-16} \\ 0 & 0 & 1.0 \end{pmatrix}.$$

Iteration 81:

$$a^{41} = (4.4e^{-5}, 2.2e^{-4}, 5.3e^{12}), \quad b^{40} = (3.6e^{4}, 9.1e^{3}, 3.8e^{-13}), \quad P^{40} = \begin{pmatrix} 1.6 & 0.40 & 4.2e^{-17} \\ 0 & 2.0 & 2.1e^{-16} \\ 0 & 0 & 2.0 \end{pmatrix}.$$

Of course, in this case, the matrices P^*, Q^*, R^* from Theorem 11, 17 are given by

$$P^* = \begin{pmatrix} 8/5 & 2/5 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad Q^* = \begin{pmatrix} 2 & 1/2 & 0 \\ 0 & 5/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad R^* = \begin{pmatrix} 4/\sqrt{5} & 1/\sqrt{5} & 0 \\ 0 & \sqrt{5} & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}.$$

Finally, to get $\bar{R}^*$ from Theorem 22, it suffices to normalize R^*.

This very simple example illustrates the different points developed in this article:

- When R does not have only positive entries, the limits of the sequences (P^n) and (Q^n) given by the Sinkhorn algorithm may be different and have more zero entries than R;
- Because new zero entries appear, the potentials (a^n) and (b^n) that are updated at each iteration of the Sinkhorn algorithm cannot converge: some of their coordinates have to tend to 0 and then some other ones need to diverge to $+\infty$ as the number of iterations increases;
- More precisely, for (i, j) on the common support of P^* and Q^* from Theorem 11, the infinitely small and high values of the two potentials are compensated. For (i, j) outside of this common support, but still in the one of R, the multiplication of the two potentials generate infinitely small values. Outside of the support of R, the multiplication of the potentials can diverge. Also, the zero entries of R prevent the sums involved in the computations of (a^n) and (b^n) to diverge: the large values of the potentials are sent to zero in the multiplication with R;
- When the problem is non-scalable, the algorithm still converges to two limits and the algorithm alternates between them. These two limits correspond to solutions of the Schrödinger problem with modified marginals, that is with modified μ or modified ν alternatively (see the iterations 80 and 81).

Going back to the context of the beginning of the section, where the upper-diagonal R models the evolution of the quantity of mRNAs corresponding to one gene in a population of cells between two timepoints, we see that the non-scalable case appears when there exists a threshold such that more cells with less mRNAs than the threshold are measured at the second timepoint than at the first one, which is incompatible with the model where the quantity of mRNAs can only increase. If we believe enough in our model, it is natural to look for a solution with modified marginals – like for instance the law $\bar{R}^*$ described in Theorem 22 – and to advocate for a bad sampling or imprecise measurements when collecting data.

If we consider that such incompatibilities between the theoretical model R and the observations μ and ν should be rare, this law $\bar{R}^*$ is a natural choice since among all the solutions of a Schrödinger problem w.r.t. R, it is the one whose marginals are the closest (in a specific entropic sense) to the experimental ones.

B Proof of Theorem 23

In this section, we prove Lemma 24 and then Theorem 23. The following classical proposition will be used in the proof of Lemma 24:

Proposition 36. *Let $R \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F})$, $\mu \in \mathcal{M}_+(\mathcal{D})$ and $\nu \in \mathcal{M}_+(\mathcal{F})$ be such that the problem $\mathrm{Sch}(R; \mu, \nu)$ admits a solution R^*. Then for all $P \in \Pi(\mu, \nu)$ such that $P \ll R$, we necessarily have $P \ll R^*$.*

Proof of Proposition 36. Let $P \in \Pi(\mu,\nu)$ be such that $P \ll R$. For all $\varepsilon > 0$, $P^\varepsilon := (1-\varepsilon)R^* + \varepsilon P$ is a competitor for $\mathrm{Sch}(R;\mu,\nu)$, so that by minimality of R^*, $H(P^\varepsilon|R) \geq H(R^*|R)$. But because $s \mapsto s\log s$ is decreasing with infinite slope near $s = 0$, this is possible for ε small only if $P \ll R^*$. □

Let us now prove Lemma 24.

Proof of Lemma 24. In Lemma 24, we are looking for necessary and sufficient conditions for $\mathrm{Sch}(R;\mu,\nu)$ to be scalable. As $\mathsf{M}(\mu) = \mathsf{M}(\nu)$ is clearly a necessary condition because of Remark 2, we assume once for all that it is true. Up to normalizing, we assume that $\mu \in \mathcal{P}(\mathcal{D})$ and $\nu \in \mathcal{P}(\mathcal{F})$.

In order to clarify what (a') and (b') mean, we start by considering the case where the reference matrix R is such that the graph $(\mathcal{D}\cup\mathcal{F}, \triangle)$ is connected. In that case, recalling that μ and ν are assumed to be probability measures, the conditions (a') and (b') are equivalent to:

(a") The measures μ and ν have full support, and for all $\emptyset \subsetneq A \subsetneq \mathcal{D}$, $\mu(A) < \nu(F_R(A))$,

(b") The measures μ and ν have full support, and for all $\emptyset \subsetneq B \subsetneq \mathcal{F}$, $\nu(B) < \mu(D_R(B))$.

Indeed, it is easy to see that in the balanced and connected case, the only subsets A of $\mathcal{D}$ for which $\mu^R(A) = \nu^R(F_R(A))$ are $A = \emptyset$ and $A = \mathcal{D}$. Similarly, the only subsets B of $\mathcal{F}$ for which $\nu^R(B) = \mu^R(D_R(B))$ are $B = \emptyset$ and $B = \mathcal{F}$.

We are going to prove Lemma 24 under this connectivity assumption. Passing to the general case is direct, up to restricting ourselves to connected components of $(\mathcal{D}\cup\mathcal{F}, \triangle)$.

We only prove (b")⇔(c'), as (a")⇔(c') is proved in the same way. The idea of the proof is to fix $\mu \in \mathcal{P}(\mathcal{D})$ of full support, that is, such that for all $i = 1,\dots,N$, $\mu_i > 0$, and to introduce the two following subsets of $\mathcal{P}(\mathcal{F})$:

$$\mathfrak{A} := \{\nu \in \mathcal{P}(\mathcal{F}) \,|\, \forall \emptyset \subsetneq B \subsetneq \mathcal{D},\ \nu(B) < \mu(D_R(B)) \text{ and } \forall j = 1,\dots,M,\ \nu_j > 0\},$$
$$\mathfrak{B} := \{\nu := \nu^{\bar{R}} \,|\, \bar{R} \sim R \text{ and } \mu^{\bar{R}} = \mu\}.$$

With these definitions, proving (b")⇔(c') exactly means proving

$$\mathfrak{A} = \mathfrak{B},$$

and this is what we will prove now. To do so, we will first show that $\mathfrak{B} \subset \mathfrak{A}$, and then that $\mathfrak{B}$ is open and closed in $\mathfrak{A}$. As $\mathfrak{A}$ is convex, and hence connected, the result will follow.

Step 1: $\mathfrak{B} \subset \mathfrak{A}$.

Let us show that $\mathfrak{B} \subset \mathfrak{A}$. To this end, let us consider $\nu \in \mathfrak{B}$, and $\bar{R}$ such that $R \in \Pi(\mu,\nu)$ and $\bar{R} \sim R$. For all $\emptyset \subsetneq B \subsetneq \mathcal{D}$ we have:

$$\nu(B) = \sum_{x_i\in\mathcal{D}}\sum_{y_j\in B}\bar{R}_{ij} = \sum_{x_i\in D_R(B)}\sum_{y_j\in B}\bar{R}_{ij} \leq \sum_{x_i\in D_R(B)}\sum_{y_j\in\mathcal{F}}\bar{R}_{ij} = \mu(D_R(B)),$$

and the equality holds only if for all $(x_i,y_j) \in D_R(B)\times B^c$, $\bar{R}_{ij} = 0$ and hence $R_{ij} = 0$. As by definition of $D_R(B)$, for all $(x_i,y_j) \in D_R(B)^c \times B$, we have $R_{ij} = 0$, an equality in the formula above would imply that $D_R(B)\times B$ is not connected to $D_R(B)^c \times B^c$ in $(\mathcal{D}\times\mathcal{F}, \triangle)$, which contradicts our connectivity assumption. We have thus a strict inequality. Finally, the full support of ν is also an consequence of the connectivity of $(\mathcal{D}\times\mathcal{F}, \triangle)$: For all $y_j \in \mathcal{F}$, let $x_i \in \mathcal{D}$ such that $R_{ij} > 0$, and hence $\bar{R}_{ij} > 0$. We have $\nu_j \geq \bar{R}_{ij} > 0$. Thus, $\mathfrak{B} \subset \mathfrak{A}$.

Step 2: $\mathfrak{B}$ is open in $\mathfrak{A}$.

It is clear by its definition that $\mathfrak{A}$ is open in $\mathcal{P}(\mathcal{F})$. Therefore, to prove that $\mathfrak{B}$ is open in $\mathfrak{A}$, it suffices to prove that $\mathfrak{B}$ is open in $\mathcal{P}(\mathcal{F})$. Let $\nu \in \mathfrak{B}$, and $\bar{R}$ be such that $\bar{R} \sim R$ and $\nu^{\bar{R}} = \nu$. We choose:

$$0 < \varepsilon < \min\{\bar{R}_{ij} \,|\, i, j \text{ s.t. } \bar{R}_{ij} > 0\}, \tag{47}$$

which is possible because $\mathcal{D}$ and $\mathcal{F}$ are finite. By convexity, it is enough to prove that for all $j \neq j'$ in $\{1, \dots, M\}$, $\nu + \varepsilon(\delta_j - \delta_{j'}) \in \mathfrak{B}$. As $(\mathcal{D} \cup \mathcal{F}, \triangle)$ is connected, we can find $j = j_0, i_1, j_1, \dots, i_p, j_p = j'$ such that

$$y_j = y_{j_0} \triangle x_{i_1} \triangle y_{j_1} \triangle \dots \triangle x_{i_p} \triangle y_{j_p} = y_{j'}.$$

Then we set

$$P := \bar{R} + \varepsilon \sum_{n=1}^{p} \Big(\delta_{i_n j_{n-1}} - \delta_{i_n j_n} \Big).$$

It is easy to check that $P \sim R$ (by the definition (47) of ε), that $\mu^P = \mu$, and that $\nu^P = \nu + \varepsilon(\delta_j - \delta_{j'})$, which therefore belongs to $\mathfrak{B}$.

Step 3: $\mathfrak{B}$ is closed in $\mathfrak{A}$, strategy of the proof.

Let us introduce the following subset of $\mathcal{P}(\mathcal{F})$:

$$\mathfrak{C} := \{\nu := \nu^{\bar{R}} \,|\, \bar{R} \ll R \text{ and } \mu^{\bar{R}} = \mu\}.$$

This set is clearly closed in $\mathcal{P}(\mathcal{F})$, so to prove that $\mathfrak{B}$ is closed in $\mathfrak{A}$, it suffices to prove that $\mathfrak{B} = \mathfrak{A} \cap \mathfrak{C}$. As we have already seen that $\mathfrak{B} \subset \mathfrak{A}$, and as clearly $\mathfrak{B} \subset \mathfrak{C}$, the only inclusion that needs to be justified is $\mathfrak{A} \cap \mathfrak{C} \subset \mathfrak{B}$.

Let us choose $\nu \in \mathfrak{A} \cap \mathfrak{C}$, and let us consider R^* the solution of $\mathrm{Sch}(R; \mu, \nu)$. We will prove by contradiction that $R^* \sim R$, and hence that $\nu \in \mathfrak{B}$.

So let us assume that $R^* \nsim R$. Once again, we choose $0 < \varepsilon < \min\{R^*_{ij} \,|\, i, j \text{ s.t. } R^*_{ij} > 0\}$. For all i, j, we write

$$x_i \blacktriangle y_j$$

whenever $R^*_{ij} > 0$. We will first prove that $(\mathcal{D} \cup \mathcal{F}, \blacktriangle)$ is connected (this is the hardest part of the proof), and then that it coincides with $(\mathcal{D} \cup \mathcal{F}, \triangle)$, which exactly means that $R^* \sim R$.

Step 4: $(\mathcal{D} \cup \mathcal{F}, \blacktriangle)$ is connected.

We call $\mathcal{D}_1 \cup \mathcal{F}_1, \dots, \mathcal{D}_p \cup \mathcal{F}_p$ the connected components of $(\mathcal{D} \cup \mathcal{F}, \blacktriangle)$. Let us assume that $p > 1$, and show that it leads to a contradiction.

First, we claim that if $p > 1$, there exist $k_1, \dots, k_l \in \{1, \dots, p\}$ a family of two by two distinct indices, $x_{i^{k_1}}, \dots, x_{i^{k_l}} \in \mathcal{D}$ and $y_{j^{k_1}}, \dots, y_{j^{k_l}} \in \mathcal{F}$ such that for all $q = 1, \dots, l$, $x_{i^{k_q}} \in \mathcal{D}_{k_q}$ and $y_{j^{k_q}} \in \mathcal{F}_{k_q}$, and with the convention $l + 1 = 1$, $y_{j^{k_q}} \triangle x_{i^{k_{q+1}}}$.

For proving this claim, we start by building a directed graph structure on $\{1, \dots, p\}$, the set of indices of the connected components of $(\mathcal{D} \cup \mathcal{F}, \blacktriangle)$. For all $k, k' \in \{1, \dots, p\}$, we write $k \rightsquigarrow k'$ whenever $k \neq k'$ and there exists $y_j \in \mathcal{F}_k$ and $x_i \in \mathcal{D}_{k'}$ such that $y_j \triangle x_i$. Of course, our claim precisely means that the directed graph $(\{1, \dots, p\}, \rightsquigarrow)$ admits a cycle. Let us prove that for all $k = 1, \dots, p$, there exists $k' \in \{1, \dots, p\}$ such that $k \rightsquigarrow k'$, which is clearly enough to conclude.

Let us consider $k \in \{1, \dots, p\}$. Because $\nu \in \mathfrak{A}$, we have $\mu(D_R(\mathcal{F}_k)) > \nu(\mathcal{F}_k)$. On the other hand, $\nu(\mathcal{F}_k) = \mu(\mathcal{D}_k)$ (as under R^*, all the mass on $\mathcal{D}_k$ is sent to $\mathcal{F}_k$ and *vice versa*), and $\mathcal{D}_k \subset D_R(\mathcal{F}_k)$ (as $R^* \ll R$). Therefore, $\mu(D_R(\mathcal{F}_k) \backslash \mathcal{D}_k) = \mu(D_R(\mathcal{F}_k)) - \mu(\mathcal{D}_k) > 0$, and we conclude that $D_R(\mathcal{F}_k) \backslash \mathcal{D}_k \neq \emptyset$. Let $x_i \in D_R(\mathcal{F}_k) \backslash \mathcal{D}_k$, and k' such that $x_i \in \mathcal{D}_{k'}$. As $x_i \in D_R(\mathcal{F}_k)$, there is $y_j \in \mathcal{F}_k$ such that $y_j \triangle x_i$. It follows that $k \rightsquigarrow k'$, and the claim is proved, and we can consider $k_1, \dots, k_l$ satisfying the properties above.

We now show that we reach a contradiction, which allows to conclude that actually, p must be equal to 1 and hence that $(\mathcal{D} \cup \mathcal{F}, \blacktriangle)$ needs to be connected.

For all $q = 1, \dots, l$, as $(\mathcal{D}_{k_q} \cup \mathcal{F}_{k_q}, \blacktriangle)$ is connected, we can find a family of indices $i^{k_q} = i_1^{k_q}, j_1^{k_q}, \dots, i_{n_q}^{k_q}, j_{n_q}^{k_q} = j^{k_q}$ such that

$$x_{i^{k_q}} = x_{i_1^{k_q}} \blacktriangle y_{j_1^{k_q}} \blacktriangle \dots \blacktriangle x_{i_{n_q}^{k_q}} \blacktriangle y_{j_{n_q}^{k_q}} = y_{j^{k_q}}$$

Now, still keeping the convention $l + 1 = 1$, we set

$$P := R^* + \varepsilon \sum_{q=1}^{l} \delta_{i^{k_{q+1}} j^{k_q}} - \varepsilon \sum_{q=1}^{l} \left(\sum_{n=1}^{n_q - 1} \left(\delta_{i_n^{k_q} j_n^{k_q}} - \delta_{i_{n+1}^{k_q} j_n^{k_q}} \right) + \delta_{i_{n_q}^{k_q} j_{n_q}^{k_q}} \right).$$

The matrix P has less zeros than R^*: by the definition (47) of ε, it has no additional zero and we have for instance $P_{i^{k_2} j^{k_1}} > 0$ and $R^*_{i^{k_2} j^{k_1}} = 0$. Moreover, $P \in \Pi(\mu, \nu)$, and the construction of the indices ensures that P has new non-zero entries w.r.t. R^* only on (x_i, y_j) such that $R_{ij} > 0$, which ensures that $P \ll R$. In virtue of Proposition 36, this contradicts the fact that R^* is the solution of $\mathrm{Sch}(R; \mu, \nu)$, and we conclude that $p = 1$.

Step 5: $(\mathcal{D} \cup \mathcal{F}, \blacktriangle) = (\mathcal{D} \cup \mathcal{F}, \triangle)$.

Our last task is to prove that whenever $x_i \triangle y_j$, for some $x_i \in \mathcal{D}$ and $y_j \in \mathcal{F}$, then we also have $x_i \blacktriangle y_j$ (the reciprocal statement follows from $R^* \ll R$). So let us consider $x_i \in \mathcal{D}$ and $y_j \in \mathcal{F}$ with $x_i \triangle y_j$, assume that we do not have $x_i \blacktriangle y_j$, and show that we reach a contradiction. As we know that $(\mathcal{D} \cup \mathcal{F}, \blacktriangle)$ is connected., we can find $i = i_1, j_1, i_2, \dots, i_p, j_p = j'$ such that

$$x_i = x_{i_1} \blacktriangle y_{j_1} \blacktriangle x_{i_2} \dots \blacktriangle x_{i_p} \blacktriangle y_{j_p} = y_j.$$

We set

$$P := R^* + \varepsilon \delta_{ij} - \varepsilon \left(\sum_{n=1}^{p-1} \left(\delta_{i_n j_n} - \delta_{i_{n+1} j_n} \right) + \delta_{i_p j_p} \right).$$

Once again, P has less zeros than R^*, which is a contradiction, and the result follows. □

We are now ready to conclude the proof of Theorem 23.

Proof of Theorem 23. It remains to show that if μ and ν are such that (a) is verified (and not (a')), then the problem is approximately scalable (once again, (b)⇒(c) is proved in the same way, and (c)⇒(a) and (c)⇒(b) are easy).

Let us first remark that the problem $\mathrm{Sch}(R; \mu^R, \nu^R)$ is obviously scalable. Let us consider $\varepsilon \in (0, 1)$. We define:

$$\mu^\varepsilon := (1 - \varepsilon)\mu + \varepsilon \mu^R \quad \text{and} \quad \nu^\varepsilon = (1 - \varepsilon)\nu + \varepsilon \nu^R.$$

The condition (a) implies that for all $A \subset \mathcal{D}$:

$$\mu^\varepsilon(A) \leq (1 - \varepsilon)\nu(F_R(A)) + \varepsilon \nu^R(F_R(A)) = \nu^\varepsilon(A).$$

with a strict inequality whenever $\mu^R(A) < \nu^R(F_R(A))$.

Let us now assume that Assumption 9 holds. In this case, in virtue of Lemma 24, the problem $\mathrm{Sch}(R; \mu^\varepsilon, \nu^\varepsilon)$ is scalable. In particular, there exists $R^\varepsilon \in \Pi(\mu^\varepsilon, \nu^\varepsilon)$ such that R^ε and R have the same support. As the family (R^ε) has value in the compact set

$$\{R' \in \mathcal{M}_+(\mathcal{D} \times \mathcal{F}) \,|\, \mathsf{M}(R') \leq \max(\mathsf{M}(\mu), \mathsf{M}(R))\},$$

we can chose one of its limit points $\bar{R}$ as $\varepsilon \to 0$. Obviously, $\bar{R} \ll R$ and $\bar{R} \in \Pi(\mu, \nu)$ so that $\mathrm{Sch}(R; \mu, \nu)$ is approximately scalable.

It remains to prove that (a)⇒(c) even when Assumption 9 does not hold. To do so, we claim that under (a), assuming Assumption 9 is not restrictive. The reason is that (a) implies Assumption 8, and hence Assumption 9 up to restricting the problem to the supports of μ and ν, as explained in Subsection 2.3.

So let us prove that (a) implies Assumption 8. We suppose that (a) holds, and we consider $\mathcal{E}$ and R^0 as defined in Assumption 8.

Let us show that $\mu \ll \mu^{R^0}$. Let $x_i \in \mathcal{D}$ be such that $\mu_i > 0$, and let us show that $\mu_i^{R^0} > 0$. By (a), $\nu(F_R(\{x_i\}) \geq \mu_i > 0$. Therefore, $F_R(\{x_i\})$ is nonempty, and there exists $y_j \in F_R(\{x_i\})$ such that $\nu_j > 0$. This pair (x_i, y_j) belongs to $\mathcal{E}$, so $R^0_{ij} > 0$, and then $\mu_i^{R^0} > 0$.

Let us show that $\nu \ll \nu^{R^0}$. Let $y_j \in \mathcal{F}$ be such that $\nu_j > 0$, and let us show that $\nu_j^{R^0} > 0$. Let us call $\mathcal{D}'$ the support of μ. By (a), $\mathsf{M}(\nu) \geq \nu(F(\mathcal{D}')) \geq \mu(\mathcal{D}') = \mathsf{M}(\mu)$. But as $\mathsf{M}(\nu) = \mathsf{M}(\mu)$, we conclude that $\nu(F(\mathcal{D}')) = \mathsf{M}(\nu)$, and so in particular that $y_j \in F(\mathcal{D}')$. So there exists $x_i \in \mathcal{D}'$ such that $R_{ij} > 0$. This pair (x_i, y_j) belongs to $\mathcal{E}$, so $R^0_{ij} > 0$, and then $\nu_j^{R^0} > 0$. □

References

[1] E. Achilles. "Implications of convergence rates in Sinkhorn balancing". In: *Linear Algebra and its Applications* 187 (1993), pp. 109–112.

[2] J-D. Benamou et al. "Iterative Bergman projections for regularized transportation problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138.

[3] A. Braides. *Gamma-convergence for Beginners*. Vol. 22. Clarendon Press, 2002.

[4] R. Brualdi. "Convex sets of non-negative matrices". In: *Canadian Journal of Mathematics* 20 (1968), pp. 144–157.

[5] G. Carlier et al. "Convergence of entropic schemes for optimal transport and gradient flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418.

[6] L. Chizat et al. "Scaling algorithms for unbalanced optimal transport problems". In: *Mathematics of Computation* 87.314 (2018), pp. 2563–2609.

[7] I. Csiszár. "I-divergence geometry of probability distributions and minimization problems". In: *The annals of probability* (1975), pp. 146–158.

[8] M. Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in neural information processing systems*. 2013, pp. 2292–2300.

[9] H. Föllmer. "Random fields and diffusion processes". In: *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. Springer, 1988, pp. 101–203.

[10] R. Fortet. "Résolution d'un système d'équations de M. Schrödinger". In: *J. Math. Pures Appl.* 19 (1940), pp. 83–105. ISSN: 0021-7824.

[11] I. Gentil, C. Léonard, and L. Ripani. "About the analogy between optimal transport and minimal entropy". In: *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6* 3 (2017), pp. 569–600.

[12] A. Hagberg, P. Swart, and D. Chult. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[13] U. Herbach et al. "Inferring gene regulatory networks from single-cell data: a mechanistic approach". In: *BMC Systems Biology* 11 (2017), p. 105.

[14] M. Idel. *A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps*. 2016. DOI: 10.48550/ARXIV.1609.06349. URL: https://arxiv.org/abs/1609.06349.

[15] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. "A new optimal transport distance on the space of finite Radon measures". In: *Advances in Differential Equations* 21.11/12 (2016), pp. 1117–1164.

[16] H. Lavenant et al. "Towards a mathematical theory of trajectory inference". In: *arXiv preprint arXiv:2102.09204* (2021).

[17] C. Léonard. "A survey of the Schrödinger problem and some of its connections with optimal transport". In: *Discrete & Continuous Dynamical Systems-A* 34.4 (2014), pp. 1533–1574.

[18] C. Léonard. "From the Schrödinger problem to the Monge–Kantorovich problem". In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920.

[19] C. Léonard. "Some properties of path measures". In: *Séminaire de Probabilités XLVI*. Springer, 2014, pp. 207–230.

[20] M. Liero, A. Mielke, and G. Savaré. "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures". In: *Inventiones mathematicae* 211.3 (2018), pp. 969–1117.

[21] T. Mikami. "Monge's problem with a quadratic cost by the zero-noise limit of h-path processes". In: *Probability theory and related fields* 129.2 (2004), pp. 245–260.

[22] M. Nutz. "Introduction to Entropic Optimal Transport". In: ().

[23] G. Peyré and M. Cuturi. "Computational optimal transport: With applications to data science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[24] L. Rüschendorf and W. Thomsen. "Note on the Schrödinger equation and I-projections". In: *Statist. Probab. Lett.* 17.5 (1993), pp. 369–375. ISSN: 0167-7152. DOI: 10.1016/0167-7152(93)90257-J. URL: https://doi.org/10.1016/0167-7152(93)90257-J.

[25] I.N. Sanov. *On the probability of large deviations of random variables*. Tech. rep. North Carolina State University. Dept. of Statistics, 1958.

[26] G. Schiebinger et al. "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming". In: *Cell* 176 (2019), 928–943 e22.

[27] E. Schrödinger. "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique". In: *Annales de l'institut Henri Poincaré*. Vol. 2. 4. 1932, pp. 269–310.

[28] E. Schrödinger. *Über die Umkehrung der naturgesetze*. Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u. Company, 1931.

[29] R. Sinkhorn. "A relationship between arbitrary positive matrices and doubly stochastic matrices". In: *The annals of mathematical statistics* 35.2 (1964), pp. 876–879.

[30] R. Sinkhorn. "Diagonal equivalence to matrices with prescribed row and column sums". In: *The American Mathematical Monthly* 74.4 (1967), pp. 402–405.

[31] E. Ventre. "Analyse, calibration et évaluation de modèles stochastiques d'expression des gènes". In: *Hal* (2022).

[32] E. Ventre. "Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics". In: *In Silico Biology* Preprint (2022), pp. 1–25.

[33] E. Ventre et al. "One model fits all: combining inference and simulation of gene regulatory networks". In: *bioRxiv* (2022).

[34] J-C. Zambrini. "Variational processes and stochastic versions of mechanics". In: *Journal of Mathematical Physics* 27.9 (1986), pp. 2307–2330.

Chapter 6

Analysis of the dynamical Schrödinger problem and application to simulated datasets.

In this chapter, we return to the dynamical Schrödinger problem (1.5), when the reference R is the path measure associated to a stochastic process modeling gene expression dynamics. We have seen in the previous chapter that at least in the discrete case, for the reference coupling R_{0T} defined by (1.3) and two empirical measures μ and ν describing the observations at two timepoints $t = 0$ and $t = T$ (in hours), the Sinkhorn algorithm always gives access to a relevant coupling:

$$R^*_{0T} = \lim_{\lambda \to \infty} \operatorname*{argmin}_P H(P|R_{0T}) + \lambda H(X_\# P|\mu) + \lambda H(Y_\# P|\nu).$$

This coupling also coincides with the solution of the problem $\mathrm{Sch}(R_{0T}; \mu^*, \nu^*)$, where μ^* and ν^* are defined by the formula (12) of Chapter 5.
In this last chapter, we aim to link this coupling R^*_{0T} to the solution of the dynamical Schrödinger problem (1.5), denoted R^*, when the reference R is the path measure associated to the bursty model (1.17). In that case, the main challenge is to derive an explicit form for the entropy relatively to such process, that we do in Section 6.2. Then, we find in Section 6.3 a formula linking the jump kernel of the reference process to the one associated to the solution of the Schrödinger problem. As in the case of SDEs for the optimal velocity [], this new kernel depends on time and space. In order to be able to interpret the jump kernel as a function of the burst rate functions k_{on} and k_{off}, we provide beforehand in Section 6.1 a general proof of the convergence of a PDMP process of the form (1.15) to a bursty process of the form (1.17). We then derive from these results a methodology for analyzing single-cell data given two experimental observations and a reference process, and we present some results on very simple *in silico* generated datasets, as a proof of concept. We emphasize that contrary to optimal-transport methods in machine learning, that are generally well-suited to high-dimensional datasets, this method is thought to be relevant for relatively small networks, of the same order than the one obtained by CARDAMOM in Chapter 4. Furthermore, it is worth noticing that the method that we develop is only suited for the moment for pair of timepoints, and not for time-course series of datasets as it was the case in Part II, preventing us from comparing the results to the ones obtained with CARDAMOM on experimental datasets. More importantly, the fact that we will perform the Schrödinger problem analysis when simplified models representing only proteins dynamics are taken as reference, the practical section will be limited to single-cell proteomic observations, instead of transcriptomic profiles as in the previous parts. We nevertheless believe that this limitation could be overcome by a scaling analysis similar to the one performed in Chapter 3 when developing CARDAMOM, for approximating the main modes of the protein levels from the observation of the mRNA levels. This work is still in progress and as such, some mathematical results will need to be refined

in the future. Most of the results presented here have been obtained with the occasional but important help of Aymeric Baradat.

We consider a generalization of the PDMP and bursty processes presented in Section 1.2 and used in Parts I and II. We fix $T > 0$ in the following.
We define a bursty process, with values in $\mathbb{R}^n$. It is defined by a time-dependent drift $F : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$, differentiable with respect to the variable in $\mathbb{R}^n$ and uniformly bounded together with its derivatives, and a time-dependent jump probability kernel $q : \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ such that for every test function $\phi : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$, the generator is of the form:

$$\mathcal{A}_{q,F}\phi(t,x) = \partial_t\phi(t,x) + \langle F(t,x), \nabla\phi(t,x)\rangle + \int (\phi(t,y) - \phi(t,x))\, q(t,x,y)dy. \tag{6.1}$$

q must be such that there exists a positive function $\lambda : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^+$ and a probability kernel $p : \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ such that for all t, x, $p(t,x,\cdot)$ is a probability measure on $\mathbb{R}^n$, and for all t, x, y, $q(t,x,y) = \lambda(t,x)p(t,x,y)$. The jump rate λ is also differentiable with respect to the variable in $\mathbb{R}^n$ and uniformly bounded together with its derivatives. $R^{q,F}$ denotes a measure, on the space $càdlàg([0,T], \mathbb{R}^n)$ associated to a bursty process with generator $\mathcal{A}_{q,F}$ defined by (6.1).

We define a PDMP process, with values in $P_E \times \mathbb{R}^n$, by a time-dependent drift $F : \mathbb{R}^+ \times P_E \times \mathbb{R}^n \to \mathbb{R}$, differentiable with respect to the variable in $\mathbb{R}^n$ and uniformly bounded together with its derivatives, and a time-dependent jump probability kernel $Q : \mathbb{R}^+ \times P_E \times P_E \times \mathbb{R}^n \to \mathbb{R}^+$ such that for every test function $\phi : \mathbb{R}^+ \times P_E \times \mathbb{R}^n \to \mathbb{R}$, the generator is of the form:

$$\mathcal{A}_{Q,F}\phi(t,e,x) = \partial_t\phi(t,e,x) + \langle F(t,e,x), \nabla\phi(t,e,x)\rangle + \sum_{e' \in P_E} \big(\phi(t,e',x) - \phi(t,e,x)\big)\, Q(t,e,e',x). \tag{6.2}$$

Q must be such that there exists a positive function $\lambda : \mathbb{R}^+ \times P_E \times \mathbb{R}^n \to \mathbb{R}^+$ and a probability kernel $p : \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ such that for all t, e, x, $p(t,e,\cdot,x)$ is a probability measure on P_E, and for all t, e, e', x, $Q(t,e,e',x) = \lambda(t,e,x)p(t,e,e',x)$. The jump rate λ is also differentiable with respect to the variable in $\mathbb{R}^n$ and uniformly bounded together with its derivatives.
$R^{Q,F}$ denotes a measure, on the space $càdlàg([0,T], P_E \times \mathbb{R}^n)$, associated to a PDMP process with generator $\mathcal{A}_{Q,F}$ defined by (6.2).

Note that in the following, we always denote by q, Q, and F the quantities described above without mentioning the specific assumptions that characterize them.

6.1 Convergence of the PDMP model to the bursty model

In order to study the Schrödinger problem when the reference is the bursty model (1.17), we need to provide beforehand a general setting for the convergence of the PDMP model (1.15) to the bursty model (1.17), which had only been stated for bursts rate functions $k_{on,i} \in C(\mathbb{R}^n, \mathbb{R})$ constant in time and scalar function $k_{off,i} \in \mathbb{R}$ in the introduction.

Indeed, as detailed in the previous section, the Schrödinger problem allows to find an optimal coupling knowing observations (at two timepoints) and a reference coupling. From the results presented in Section 1.1.2, this optimal coupling could be related to an optimal stochastic process, that we would like to relate to its characteristics (for example its kernel Q and its drift F if it is a PDMP process of the form (6.2)).
If we were able to study the dynamical Schrödinger problem for the PDMP model (1.15), we could deduce from Q^* the optimal burst rate functions $k^*_{on,i}$ and $k^*_{off,i}$ for every gene i, and use a parametric form for the burst rate functions similar to the one described in the formula (2) of

Chapter 3 in order to deduce an optimal GRN. However, since the promoters are not observed, solving the Schrödinger problem for this model is not possible, at least without considering that the temporal marginals only describes a subset of the variables describing the process, which would complicate the analyzes. We need for a model which takes only into account the expression levels.
We are then going to consider the bursty model (1.17), which is completely characterized by protein levels, and thus more likely to be used for the Schrödinger problem (even if, as mentioned in the introduction, we would like to be able to deal with mRNA levels, but this leads to too complex model at this stage). The bursty model is characterized by a deterministic drift and a transition kernel q, but the latter has not a parametric form driven by mechanistic assumptions. The form used in Chapter 3 results from the convergence of the PDMP model towards the bursty model for scalar $k_{off,i}$ functions. Thus, if the functions $k_{off,i}$ are not scalar, it is not clear at the moment what would be the form of the kernel q of the limit process. We then need a more general convergence result in order to link the non-parametric characteristics a an optimal bursty process to the parametric ones of an associated PDMP process.
Moreover, we would like to be sure that solving the Schrödinger problem when the reference measure is the one of a PDMP process is equivalent to solve the Schrödinger problem when the reference measure is the one of the limit bursty process. That is why we are going to study the convergence of the PDMP model (6.2) to the bursty model (6.1), and the Γ-convergence of the entropies relatively to these two processes.

We consider the space of promoters $P_E := \{0,1\}^n$, where n is the number of genes of interest, and that the kernel Q is such that $\forall e, e' \in P_E$, $\forall t > 0$:

$$\begin{cases} Q^\varepsilon(t,e,e',x) = k_{on,i}(t,x)\mathbb{1}_{e_i'-e_i=1} + \frac{1}{\varepsilon}k_{off,i}(t,x)\mathbb{1}_{e_i-e_i'=1} & \text{if } \exists i : \forall j \neq i, e_j = e_j' \text{ and } e_i \neq e_i', \\ Q^\varepsilon(t,e,e',x) = 0 & \text{if not,} \end{cases} \tag{6.3}$$

where for all i, $k_{on,i}$ and $k_{off,i}$ are two functions differentiable with respect to their second variables and uniformly bounded together with their derivatives. We also assume that for all gene i:

$$\forall x, e \in \mathbb{R}^n \times P_E : F_i^\varepsilon(t,e,x) = F_{0,i}(t,x)\mathbb{1}_{e_i=0} + \frac{1}{\varepsilon}F_{1,i}(t,x)\mathbb{1}_{e_i=1}, \tag{6.4}$$

with $F_{0,i}, F_{1,i}$ two functions differentiable with respect to their second variables and uniformly bounded together with their derivatives.
We are interested in the case $\varepsilon \ll 1$. We are going to prove that when $\varepsilon \to 0$, the PDMP model (6.2) verifying (6.3) and (6.4) converges to a bursty process of the form (6.1). We emphasize that the aim of this analysis is to precise the form of the jump kernel q appearing in the convergence, in order to be able to relate an optimal kernel q^*, appearing in the solution of the dynamical Schrödinger problem, to the parametric functions $k_{on,i}$ and $k_{off,i}$.

6.1.1 Convergence in law

The following theorem is an extension to the multidimensional case of the Theorem 6.1 of Crudu et al. [20]. We can also simplify some steps in comparison with this reference because the creation and degradation of proteins are deterministic knowing the promoters states.

Theorem 11. *Let us consider $(R^{Q^\varepsilon,F^\varepsilon})_\varepsilon$ a sequence of measures characterizing PDMP processes of the form $(E_t^\varepsilon, X_t^\varepsilon)_{t\in[0,T]}$ for every ε, with generator defined by (6.2) and $(Q^\varepsilon, F^\varepsilon)$ verifying (6.3) and (6.4). Moreover, we assume that:*

- *For all i, $E_{i\,0}^\varepsilon \to 0$ in law as $\varepsilon \to 0$;*

- $X_0^\varepsilon \to X_0$ *in law as* $\varepsilon \to 0$, *and for all* i: $\mathbb{E}(X_{i\,0}^\varepsilon) < \infty$;
- *There exists* α *such that* $\forall t, x \in \mathbb{R}^+ \times \mathbb{R}^n$ *and for all* i: $k_{off,i}(t,x) \geq \alpha$.

Then $(E_t^\varepsilon, X_t^\varepsilon)_{t\in[0,T]}$ *converges in law to the bursty process* $(0, X_t)_{t\in[0,T]}$ *where the path measure of* $(X_t)_{t\in[0,T]}$ *is of the form* R^{q,F_0}, *with generator defined by* (6.1). *The drift* F_0 *and the kernel* q *are defined for all* $t > 0$, $x \in \mathbb{R}^n$ *and for all* i *by:*

$$q(t, x, \psi_i(t, u, x)) = k_{on,i}(t, x) P_u^i k_{off,i}(t, x), \tag{6.5}$$

where ψ_i *is the flow associated to the drift* $F_{1,i}$, *i.e for all* $t, u > 0$ *and* $x \in \mathbb{R}^n$*:*

$$\partial_u \psi_i(t, u, x)) = F_{1,i}(t, x) \partial_{x_i} \psi_i(t, u, x),$$

and for all i, P_u^i *is a semigroup defined for every test function* $\phi : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ *by:*

$$P_u^i \phi(t, x) = \phi(t, \psi_i(t, u, x)) \exp\left(- \int_0^u k_{off,i}(t, \psi_i(t, s, x)) \mathrm{d}s \right). \tag{6.6}$$

If (t, x, y) *are such that there does not exist any gene* i *and* $u > 0$ *such that* $y = \psi_i(t, u, x)$, *then* $q(t, x, y) = 0$.

Proof. We denote $\mathcal{A}_{Q^\varepsilon, F^\varepsilon}$ the generator defined by (6.2) when the jump kernel Q^ε and the drift F^ε are defined by the relations (6.3) and (6.4). We first justify for every gene i the tightness of E_i^ε and X_i^ε. The aim is to use then the Prokhorov theorem, which ensures that the sequence $((E_t^\varepsilon, X_t^\varepsilon)_{t\in[0,T]})_\varepsilon$ has a weak subsequential limit. The core of the proof consists then in proving the convergence of the generator $\mathcal{A}_{Q^\varepsilon, F^\varepsilon}$ to $\mathcal{A}_{q,F_0}$ with q defined by (6.5), which is performed at Step 2. This allows to show that any limit is the solution of the martingale problem associated to the generator $\mathcal{A}_{q,F_0}$. By unicity of the martingale problem, we will finally conclude on the convergence of the PDMP process to the bursty process.

Step 1: The sequence $((E_t^\varepsilon, X_t^\varepsilon)_{t\in[0,T]})_\varepsilon$ has a weak subsequential limit.

First, we characterize the limit of the stochastic process $(E_{i\,t}^\varepsilon)_t$ characterizing promoters i dynamics when $\varepsilon \to 0$.
We consider the test function for all i, $t \in \mathbb{R}^+$, $e, x \in P_E \times \mathbb{R}^n$:

$$\begin{cases} \phi^i(t, e, x) = 1 & \text{if } e_i = 1 \text{ and } \forall j \neq i : \ e_j = 0, \\ \phi^i(t, e, x) = 0 & \text{if not.} \end{cases}$$

Then, noting that terms with $\partial_t \phi_i, \nabla \phi_i$ are cancelled here, we have:

$$\mathcal{A}_{Q^\varepsilon, F^\varepsilon} \phi^i(t, e, x) = k_{on,i}(t, x) \mathbb{1}_{e_i=0} - \frac{k_{off,i}(t, x)}{\varepsilon} \mathbb{1}_{e_i=1}.$$

Moreover, using the classical martingale property (1.2) associated to the PDMP process with generator (6.2), applied to the function ϕ^i, we know that:

$$M^\varepsilon(t) = \mathbb{1}_{E_{i\,t}^\varepsilon=1} - \mathbb{1}_{E_{i\,0}^\varepsilon=1} - \int_0^t \left(k_{on,i}(u, X_u^\varepsilon) \mathbb{1}_{E_{i\,u}^\varepsilon=0} - \frac{k_{off,i}(u, X_u^\varepsilon)}{\varepsilon} \mathbb{1}_{E_{i\,u}^\varepsilon=1} \right) du,$$

is a martingale. Thus, using the fact that $\mathbb{E}(M\varepsilon(t)) = 0$, we obtain that for every $T > 0$ and for all $t \in [0, T]$:

$$\frac{\alpha}{\varepsilon} \mathbb{E}\left(\int_0^t \mathbb{1}_{E_{i\,u}^\varepsilon=1} du \right) \leq ||k_{on,i}||_\infty T + 1, \tag{6.7}$$

We obtain that for all i:

$$\mathbb{E}\left(\int_0^t \mathbb{1}_{E^\varepsilon_{i\,u}=1} du\right) \underset{\varepsilon\to 0}{\to} 0,$$

from which the tightness of the sequence of processes $((E^\varepsilon_{i\,t})_t)_\varepsilon$ in $L^1([0,T];\{0,1\})$ follows.

The tightness of the sequence of processes $(X^\varepsilon_{i\,t})_t)_\varepsilon$ requires a little more work. Still adapting ideas of Crudu et al. [20], we are going to show that we can control the bounded variations norm on $[0,T]$ of the random process $(X^\varepsilon_{i\,t})_t$.
We denote $BV(0,T)$ the space of functions of bounded variations on $[0,T]$. Its norm is defined for all $f \in BV(0,T)$ by:

$$||f||_{BV(0,T)} = ||f||_{L^1([0,T])} + \sup_{\sigma=(t_r)_r} \Big\{ \sum_i |f(t_{r+1}) - f(t_r)| \Big\},$$

where $(t_r)_r$ denotes a finite subdivision of $[0,T]$. As for any $K>0$, the set $\{f \in BV(0,T), ||f||_{BV(0,T)} \leq K\}$ is relatively compact in $L^1([0,T];\mathbb{R})$ (see for example in Giusti et al. [33]), it is enough to prove that for all $\delta>0$, there exists $K^\delta>0$, such that for all ε:

$$\mathbb{P}\left(||X^\varepsilon_i||_{BV(0,T)} > K^\delta\right) \leq \delta. \tag{6.8}$$

For all gene i, we can write at every time $t \in [0,T]$:

$$X^\varepsilon_{i\,t} = X^\varepsilon_{i\,0} + \int_0^t \left(F_{0,i}(s, X^\varepsilon_{i\,s})\mathbb{1}_{E^\varepsilon_{i\,s}=0} + \frac{1}{\varepsilon}F_{1,i}(s, X^\varepsilon_{i\,s})\mathbb{1}_{E^\varepsilon_{i\,s}=1}\right) \mathrm{d}s.$$

Then, for any subdivision $\sigma : 0 = t_0, \cdots, t_N = T$, we have:

$$\begin{aligned}\sum_{r=0}^{N-1} |X^\varepsilon_{i\,t_{r+1}} - X^\varepsilon_{i\,t_r}| &\leq \sum_{r=0}^{N-1} \int_{t_r}^{t_{r+1}} \left|F_{0,i}(s, X^\varepsilon_{i\,s})\mathbb{1}_{E^\varepsilon_{i\,s}=0} + \frac{1}{\varepsilon}F_{1,i}(s, X^\varepsilon_{i\,s})\mathbb{1}_{E^\varepsilon_{i\,s}=1}\right| \mathrm{d}s \\ &= \int_0^T \left|F_{0,i}(s, X^\varepsilon_{i\,s})\mathbb{1}_{E^\varepsilon_{i\,s}=0}\right| ds + \int_0^T \left|\frac{1}{\varepsilon}F_{1,i}(s, X^\varepsilon_{i\,s})\mathbb{1}_{E^\varepsilon_{i\,s}=1}\right| \mathrm{d}s.\end{aligned}$$

Thus, as the right-hand side term does not depend on the subdivision, we can pass to the sup and take the expected value to see that, using the relation (6.7):

$$\mathbb{E}\left(\sup_{\sigma=(t_r)_r} \Big\{ \sum_{r=0}^{N-1} |X^\varepsilon_{i\,t_{r+1}} - X^\varepsilon_{i\,t_r}| \Big\}\right) \leq ||F_{0,i}||_\infty T + ||F_{1,i}||_\infty \frac{\beta T + 1}{\alpha}.$$

Moreover, using the same upper bound as previously we have:

$$\mathbb{E}\left(\int_0^T |X^\varepsilon_{i\,t}| dt\right) \leq T\mathbb{E}(X^\varepsilon_{i\,0}) + ||F_{0,i}||_\infty T^2 + ||F_{1,i}||_\infty \frac{\beta T^2 + T}{\alpha}.$$

Thus, the expected value $\mathbb{E}\left(||X^\varepsilon_i||_{BV(0,T)}\right)$ is uniformly bounded, and we deduce the relation (6.8).

The tightness for all i in $L^1([0,T];\{0,1\}) \times L^1([0,T];\mathbb{R})$ of the sequence of processes $((E^\varepsilon_{i\,t}, X^\varepsilon_{i\,t})_{t\in[0,T]})_\varepsilon$ allows to apply the Prokhorov Theorem, which ensures the existence of a subsequence $(E^{\varepsilon_k}_t, X^{\varepsilon_k}_t)$ converging weakly in $L^1([0,T];\{0,1\}) \times L^1([0,T];\mathbb{R})$ as $\varepsilon \to 0$.
As we have seen that for all $t \in [0,T]$, $\lim_{\varepsilon\to 0} \mathbb{E}\left(\int_0^t \mathbb{1}_{E^\varepsilon_{i\,u}=1} du\right) = 0$, the limit is necessarily of the form $(0, X_t)$ for almost all $t \in [0,T]$. In order to simplify notations, the subsequence is still denoted $(E^\varepsilon_t, X^\varepsilon_t)_\varepsilon$.

Step 2: A limit formulation for the generator of the PDMP process.

Reasoning as previously for showing the relation (6.7), we define for all i, j the test function:

$$\begin{cases} \phi^{ij}(t,e,x) = 1 & \text{if } e_i = 1 \text{ and } e_j = 1 \text{ and } \forall k \neq i,j : \ e_k = 0, \\ \phi^{ij}(t,e,x) = 0 & \text{if not.} \end{cases}$$

We know that:

$$M(t) = \mathbb{1}_{E^\varepsilon_{i\,t}=1}\mathbb{1}_{E^\varepsilon_{j\,t}=1} - \mathbb{1}_{E^\varepsilon_{i\,0}=1}\mathbb{1}_{E^\varepsilon_{j\,0}=1} - \int_0^t \big(k_{on,i}(u,X^\varepsilon_u)\mathbb{1}_{E^\varepsilon_{i\,u}=0}\mathbb{1}_{E^\varepsilon_{j\,u}=1} - k_{on,j}(u,X^\varepsilon_u)\mathbb{1}_{E^\varepsilon_{i\,u}=1}\mathbb{1}_{E^\varepsilon_{j\,u}=0} - \frac{k_{off,i}(u,X^\varepsilon_u)+k_{off,j}(u,X^\varepsilon_u)}{\varepsilon}\mathbb{1}_{E^\varepsilon_{i\,u}=1}\mathbb{1}_{E^\varepsilon_{j\,u}=1}\big)du,$$

is a martingale. Using the fact that $E^\varepsilon_{i\,t} \to 0$ for almost all $t \in [0,T]$, we obtain that for all i, j and all $t > 0$:

$$\frac{1}{\varepsilon}\mathbb{E}\left(\int_0^t \mathbb{1}_{E^\varepsilon_{i\,u}=1}\mathbb{1}_{E^\varepsilon_{j\,u}=1}du\right) \underset{\varepsilon\to 0}{\to} 0.$$

We are know going to prove a simple lemma which details useful relations satisfied by the semigroup P^i_u:

Lemma 12. *The semigroup defined by the formula 6.6 verifies the three following relations:*

1. $\partial_u P^i_u\phi(t,x) = F_{1,i}(t,x)\partial_{x_i}(P^i_u\phi(t,x)) - k_{off,i}P^i_u\phi(t,x)$,
2. $\int_0^\infty P^i_u k_{off,i}(t,x)du = 1$,
3. $P^i_u(k_{off,i}\phi)(t,x) = P^i_u k_{off,i}(t,x) \times \phi(t,\psi_i(u,x))$.

Proof. The points 2 and 3 are obvious due to the exponential form. The first point 1 is well known but is slightly less obvious, so we detail the small justification here. On one side we have for all u, t, x:

$$\partial_u P^i_u\phi(t,x) = \big(\partial_u\psi_i(t,u,x)\nabla\phi(t,\psi_i(t,u,x)) - k_{off,i}(t,\psi_i(t,u,x))\phi(t,\psi_i(t,u,x))\big)\exp - \left(\int_0^u k_{off,i}(t,\psi_i(t,s,x))\mathrm{d}s\right).$$

On the other side:

$$\partial_{x_i} P^i_u\phi(t,x) = \big(\partial_{x_i}\psi_i(t,u,x)\nabla\phi(t,\psi_i(t,u,x)) - \int_0^u \partial_{x_i}(k_{off,i}(t,\psi_i(t,s,x)))\mathrm{d}s \times \phi(t,\psi_i(t,u,x))\big)\exp - \left(\int_0^u k_{off,i}(t,\psi_i(t,s,x))\mathrm{d}s\right).$$

Using the relation $\partial_s k_{off,i}(t,\psi_i(t,s,x)) = F_{1,i}(t,x)\partial_{x_i}k_{off,i}(t,\psi_i(t,s,x))$, we obtain:

$$\begin{aligned}\int_0^u \partial_{x_i}(k_{off,i}(t,\psi_i(t,s,x)))\mathrm{d}s &= \frac{1}{F_{1,i}(t,x)}\int_0^u \partial_s k_{off,i}(t,\psi_i(t,s,x)), \\ &= \frac{k_{off,i}(t,\psi_i(t,u,x)) - k_{off,i}(t,x)}{F_{1,i}(t,x)}.\end{aligned}$$

We finally obtain the equation 1. □

We now define a test function ϕ by, for all $s, x \in \mathbb{R}^+ \times \mathbb{R}^n$:

$$\phi^f(t,e,x) = \sum_{i=1}^n \left[\int_0^\infty P^i_u(k_{off,i}f)(t,x)du\right]\mathbb{1}_{e_i=1} + f(t,x)\mathbb{1}_{e=0}, \tag{6.9}$$

where f is such that for all t, $f(t,\cdot) \in C_b^1(\mathbb{R}^n)$. We emphasize that here $e = 0$ means that for the state $e \in P_E$, for all gene i, we have $e_i = 0$.
This test function satisfies for all t, x and e such that $e_i = 1$ and $\forall j \neq i$, $e_j = 0$:

$$\begin{aligned}
\mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(t,e,x) &= \int_0^\infty \partial_t(P_u^i(k_{off,i}f)(t,x))du + \frac{F_{1,i}}{\varepsilon}(t,x)\int_0^\infty \partial_{x_i}(P_u^i(k_{off,i}f)(t,x))du + \\
&\quad \frac{k_{off,i}}{\varepsilon}\big(f - \int_0^\infty P_u^i(k_{off,i}f)\big)(t,x)du + \sum_{j\neq i}\int_0^\infty \big(k_{on,j}P_u^j(k_{off,j}f)\big)(t,x)du, \\
&= \int_0^\infty \partial_t(P_u^i(k_{off,i}f)(t,x))du + \sum_{j\neq i}\int_0^\infty \big(k_{on,j}P_u^j(k_{off,j}f)\big)(t,x)du + \\
&\quad \frac{1}{\varepsilon}\big[\int_0^\infty \partial_u P_u^i(k_{off,i}f)(t,x)du + k_{off,i}f(t,x)\big], \\
&= \int_0^\infty \partial_t(P_u^i(k_{off,i}f)(t,x))du + \sum_{j\neq i}\int_0^\infty \big(k_{on,j}P_u^j(k_{off,j}f)\big)(t,x)du.
\end{aligned}$$

where we used the equation 1 defining P_u^i on the test function $k_{off,i}f$ from the first to the second line, and the fact that $\forall t,x : \int_0^\infty \partial_u P_u^i(k_{off,i}f)(t,x)du = -k_{off,i}f(t,x)$ from the second to the third line.

Recalling that the functions $k_{on,i}$ and $k_{off,i}$ are uniformly bounded together with their derivatives, we have then shown that the value of the generator applied on ϕ^f is in $O(1)$ when there is only one $e_i \neq 0$. It is easy to see that when there are two (or more) $e_i, e_j > 0$, with $i \neq j$, the generator applied on ϕ^f is in $O(\frac{1}{\varepsilon})$.

We also recall that for all i,j,t: $\lim_{\varepsilon\to 0}\mathbb{E}\left(\int_0^t \mathbb{1}_{E^\varepsilon_{i\,u}=1}du\right) = 0$ and $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left(\int_0^t \mathbb{1}_{E^\varepsilon_i E^\varepsilon_{j\,u}=1}du\right) = 0$.

We then obtain that for all t:

$$\forall e \in P_E, \sum_{i=1}^n e_i \geq 1 : \lim_{\varepsilon\to 0}\mathbb{E}\left(\int_0^t |\mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(u,E_u^\varepsilon,X_u^\varepsilon)|\mathbb{1}_{E_u^\varepsilon=e}\right)du = 0. \tag{6.10}$$

We also observe that:

$$\begin{aligned}
\mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(t,0,x) &= \partial_t f(t,x) + \langle F_0(t,x), \nabla f(t,x)\rangle + \sum_{i=1}^n \big(k_{on,i}(t,x)(\int_0^\infty P_u^i(k_{off,i}f)(t,x)du - f(t,x))\big), \\
&= \partial_t f(t,x) + \langle F_0(t,x), \nabla f(t,x)\rangle + \sum_{i=1}^n \int_0^\infty \big(P_u^i(k_{off,i}f)(t,x) - f(t,x)\big)\, k_{on,i}P_u^i k_{off,i}(t,x)du, \\
&= \mathcal{A}_{q,F_0}f(t,x),
\end{aligned}$$

where $\mathcal{A}_{q,F_0}$ is the generator defined in (6.1) and q is the jump kernel predicted by the theorem. The passage from the first to the second line comes from the hypothesis that $\int_0^\infty P_u^i(k_{off,i})(t,x) = 1$, and from the second to the third line from the relation 3 of Lemma 12.

We have then for the subsequence $(E_t^\varepsilon, X_t^\varepsilon)_\varepsilon$ that for all $t \in [0,T]$:

$$\lim_{\varepsilon\to 0}\mathbb{E}\left(\int_0^t \left|\mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(u,E_u^\varepsilon,X_u^\varepsilon) - \mathcal{A}_{q,F_0}f(u,X_u^\varepsilon)\right| \mathrm{d}u\right) = 0,$$

Step 3: Any subsequence that converges almost-surely is the unique solution of a martingale problem.

The Skorokhod representation Theorem implies the existence of a subsequence $(E_t^{\varepsilon_k}, X_t^{\varepsilon_k})$ which converges to $(0, X_t)$ almost surely, for almost all $t \in [0,T]$. Once again, the subsequence is still denoted $(E_t^\varepsilon, X_t^\varepsilon)_\varepsilon$.

Let us consider $t \in [0,T]$, a subdivision $0 = t_0 \leq \cdots \leq t_N = t$ and $\psi \in C_b(\mathbb{R}^{n\times N})$. The fact that f is bounded ensures that the function ϕ^f is bounded. Thus, it is clear that as ψ is also bounded, using the dominated convergence theorem:

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\left[\phi^f(t,E_t^\varepsilon,X_t^\varepsilon) - \phi^f(0,E_0^\varepsilon,X_0^\varepsilon)\right]\psi(X_{t_0}^\varepsilon,\cdots,X_{t_N}^\varepsilon)\right) = \mathbb{E}\left(\left[f(t,X_t) - f(0,X_0)\right]\psi(X_{t_0}^\varepsilon,\cdots,X_{t_N}^\varepsilon)\right).$$

Moreover, from (6.10) we have, as ψ is bounded:

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\left[\int_0^t \sum_{e\neq 0} \mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(u,E_u^\varepsilon,X_u^\varepsilon)\mathbb{1}_{E_u^\varepsilon=e}du\right]\psi(X_{t_0}^\varepsilon,\cdots,X_{t_N}^\varepsilon)\right) = 0.$$

Finally, using the fact that $\mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(t,0,X_t^\varepsilon) \to \mathcal{A}_{q,F_0}f(t,X_t)$ almost surely, for almost all $t \in [0,T]$, and that $\int_0^t \mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(t,0,X_t^\varepsilon)\mathrm{d}u$ is bounded for all $t \in [0,T]$, we can still use the dominated convergence theorem and conclude that:

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\left[\phi^f(t,E_t^\varepsilon,X_t^\varepsilon) - \phi^f(0,E_0^\varepsilon,X_0^\varepsilon) - \int_0^t \mathcal{A}_{Q^\varepsilon,F^\varepsilon}\phi^f(u,E_u^\varepsilon,X_u^\varepsilon)\mathrm{d}u\right]\psi(X_{t_0}^\varepsilon,\cdots,X_{t_N}^\varepsilon)\right) =$$
$$\mathbb{E}\left(\left[f(t,X_t) - f(0,X_0) - \int_0^t \mathcal{A}_{q,F_0}f(u,X_u)\mathrm{d}u\right]\psi(X_{t_0}^\varepsilon,\cdots,X_{t_N}^\varepsilon)\right).$$

As the term of the left-hand side is equal to 0, thanks to the martingale property of the PDMP process, we obtain:

$$\mathbb{E}\left(\left[f(t,X_t) - f(0,X_0) - \int_0^t \mathcal{A}_{q,F_0}f(u,X_u)\mathrm{d}u\right]\psi(X_{t_0}^\varepsilon,\cdots,X_{t_N}^\varepsilon)\right) = 0.$$

The martingale characterization of the bursty processes has been shown in Crudu et al. [20], Theorem 2.5, stating that the law of the bursty process determined by a generator of the form (6.1) with a drift F_0 and a jump kernel q defined by the relation (6.5), is the unique solution of the martingale problem associated to the generator $\mathcal{A}_{q,F_0}$.
We can thus conclude, by unicity of the solution of the martingale problem, that the limit $(X_t)_{t\in[0,T]}$ is a bursty process whose generator is given by the formula (6.1), with a drift F_0 and a jump kernel defined by the relation (6.5). □

When for all i, the $k_{off,i}$ are scalar functions, we recover the convergence that we used in particular in Part II of the manuscript:

Corollary 13. *In the particular case where for all i, the $k_{off,i}$ are scalar functions and that $F_{0,i}(x) = -d_i x_i$ and $F_{1,i}(x) = s_i$, where d_i and s_i are the degradation and creation rates defined in Section 1.2, we recover the convergence used throughout the manuscript. The kernel q can be defined for all t, x, y by:*

$$\begin{cases} q(t,x,y) = k_{on,i}(t,x)\frac{k_{off,i}}{s_i}e^{\frac{k_{off,i}}{s_i}h} & \text{if} \exists i : \forall j \neq i,\ x_j = y_j \text{ and} \exists h > 0,\ y_i = x_i + h, \\ q(t,x,y) = 0 & \text{if not.} \end{cases} \tag{6.11}$$

Proof. In that case, the flow ψ_i is simply defined for all t, x by:

$$\psi_i(t,x) = s_i t + x.$$

For all test function f, the generator $\mathcal{A}_{q,F_0}$ of the limit process is then, by Theorem 11:

$$\forall t,x:\ \mathcal{A}_{q,F_0}f(t,x)=\partial_t f(t,x)+\langle F_0(t,x),\nabla f(t,x)\rangle+\int_0^\infty (f(t,x+su)-f(t,x))k_{off,i}e^{-k_{off,i}u}du.$$

A simple substitution $y\leftarrow x+su$ in the integral allow to obtain the modified q of the corollary. □

6.2 Relative entropies and Gamma-convergence

In this section, we build the relative entropy, defined by the formula (2) in Chapter 1, relatively to the processes of interest. We provide an explicit formula for the entropy relatively to a bursty processes defined by (6.1). We then expose the formula defining the entropy relatively to a PDMP process of the form (6.2) (but skipping the proofs that are still under construction) and how these formula allow to recover the Γ-convergence of the entropy relatively to a PDMP process to the entropy relatively to a bursty process. The aim is to show that studying the Schrödinger problem when the reference is a bursty process is equivalent to studying the Schrödinger problem when the reference is a PDMP process, in the limit $\varepsilon\to 0$.

For simplifying the notation, we consider that the time of observations are simply 0 and 1 (then the stochastic processes are defined on $[0,1]$).

6.2.1 Entropy of the bursty model

We first show that when working with an entropy minimization problem, we can equivalently consider a pure jump process defined by a measure $R^{\tilde{q}}$ instead of a bursty process of the form (6.1) defined by a measure $R^{q,F}$, where q and $\tilde{q}$ are linked by a specific relation. This will allow us to use later important results presented by Leonard in [52] about Girsanov theory.

Equivalence of the bursty process with a pure jump process

We consider the application $\xi:\ \mathbb{R}^+\times[0,1]\times\mathbb{R}^n\to\mathbb{R}^n$ such that for all $(s,t,x)\in[0,1]\times[0,1]\times\mathbb{R}^n$:

$$\begin{cases}\partial_1\xi(s,t,x) &= F(s,\xi(s,t,x)),\\ \xi(t,t,x) &= x.\end{cases}\tag{6.12}$$

The fact that F is lipschitz ensures that the application is well defined, and satisfies for all t_1, t_2, t_3 a semi-group property, which means:

$$\xi(t_3,t_2,\xi(t_2,t_1,x))=\xi(t_3,t_1,x).$$

We have then for all $0\le t\le 1$:

$$\xi(t,0,\xi(0,t,x))=\xi(t,t,x)=x.$$

We define:

$$\forall x\in\mathbb{R}^n,\ (\xi(0,t,\cdot)_{\#}\rho(\cdot))(x)=\rho(\xi(0,t,x)),\tag{6.13}$$

$$\forall\phi\in C_c^\infty(\mathbb{R}^n),\ \mathbb{E}_{\xi_{\#}R^{q,F}}\left[\phi(X_t)\right]=\mathbb{E}_{R^{q,F}}\left[\phi(\xi(0,t,X_t))\right].\tag{6.14}$$

For all test function $\phi\in C_c^\infty(\mathbb{R}^n)$ and $t>0$, we denote $\psi(t,x)=\phi(\xi(0,t,x))$. By classical results of transport equation theory (see for example [48]) we know that:

$$\forall(t,x)\in[0,1]\times\mathbb{R}^n:\ \partial_t\psi(t,x)+\langle F(t,x),\nabla\psi(t,x)\rangle=0.\tag{6.15}$$

Then we obtain for all $t \in [0,1]$:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{\xi_{\#}R^{q,F}}\left[\phi(X_t^x)\right] &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{R^{q,F}}\left[\psi(t,X_t^x)\right], \\
&= \mathbb{E}_{R^{q,F}}\left[\partial_t\psi(t,X_t) + \langle F(\psi(t,X_t)), \nabla\psi(t,X_t)\rangle + \int\{\psi(t,y)-\psi(t,X_t)\}q(t,X_t,y)\mathrm{d}y\right], \\
&= \mathbb{E}_{R^{q,F}}\left[\int\{\phi(\xi(0,t,y))-\phi(\xi(0,t,X_t))\}q(t,X_t,y)\mathrm{d}y\right], \\
&= \mathbb{E}_{R^{q,F}}\left[\int\{\phi(z)-\phi(\xi(0,t,X_t))\}q(t,\xi(t,0,\xi(0,t,X_t)),\xi(t,0,z))|detJ_{\zeta_t}(z)|dz\right], \\
&= \mathbb{E}_{\xi_{\#}R^{q,F}}\left[\int\{\phi(z)-\phi(X_t)\}q(t,\xi(t,0,X_t),\xi(t,0,z))|detJ_{\zeta_t}(z)|dz\right].
\end{aligned}$$

where we denote $J_{\zeta_t}(z)$ the Jacobian matrix of the application $\zeta_t : z \to \xi(t,0,z)$. The second line comes from the well known Hille-Yosida theorem and the third line comes from the relation (6.15). Then, we see (formally) that the process $\xi_{\#}R^{q,F}$ is a pure jump process driven by the transition kernel $\tilde{q}(t,x,z) = q(t,\xi(t,0,x),\xi(t,0,z))|detJ_{\zeta_t}(z)|$, for all $(t,x,z) \in \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n$.

Moreover, we can show that for any kernels $q, \bar{q}$ and a drift F, we have the relation:

$$H(R^{q,F}|R^{\bar{q},F}) = H(\xi_{\#}R^{q,F}|\xi_{\#}R^{\bar{q},F}). \tag{6.16}$$

This is justified by Lemma 6 which ensures that for any measurable function $\phi : càdlàg([0,1];\mathbb{R}^n) \to càdlàg([0,1];\mathbb{R}^n)$ (seen as a random variable),

$$H(R^{q,F}|R^{\bar{q},F}) \geq H(\phi_{\#}R^{q,F}|\phi_{\#}R^{\bar{q},F}).$$

Thus, using the fact that $\zeta(t,\cdot)$ is a bijection for any t, we can apply two times the lemma to $\phi = \zeta$ and $\phi = \zeta^{-1}$ to obtain:

$$H(R^{q,F}|R^{\bar{q},F}) \geq H(\xi_{\#}R^{q,F}|\xi_{\#}R^{\bar{q},F}) \geq H(R^{q,F}|R^{\bar{q},F}).$$

For proving general results about entropy minimization problems, we will then consider in the following pure jump processes with time-dependent transition kernel $\tilde{q}$ instead of bursty processes of the form (6.1) with kernel q and drift F. For simplicity, we keep the notation q rather than $\tilde{q}$ when considering the pure jump process associated to a bursty process, and we denote its path measure R^q.

Entropy of the bursty model

In the following we consider that $a : [0,1] \times \mathbb{R}^n \times \mathbb{R}^n, \to \mathbb{R}$ is a function which is measurable and bounded. The following proposition is a simplified version of Theorems 2.6 and 2.9 of Leonard [51].

Proposition 14. *Let us consider $R^{q,F}$ and $R^{\bar{q},F}$ the laws of two bursty processes with the same drift F and kernels q and $\bar{q}$, respectively. We also assume that this process have the same initial conditions. We have then:*

$$H(R^{q,F}|R^{\bar{q},F}) = \mathbb{E}_{R^{q,F}}\left(\int_0^1\int h\left(\frac{q}{\bar{q}}\right)\bar{q}(s,X_s,y)\mathrm{d}y\mathrm{d}s\right), \tag{6.17}$$

where $\forall a,\ h(a) = a\ln a + 1 - a$.
Conversely, for any $P \in \mathcal{P}(C([0,T],\mathbb{R}^n))$ such that $H(P|R^{\bar{q}}) < \infty$, there exists a jump kernel q such that $P \sim R^{q,F}$.

Remark 15. As the relative entropy between bursty processes is equal to the one between the associated modified jump processes, this proposition falls within the framework of the theorems of Leonard [51]. We are nevertheless going to provide a formal proof of the formula (6.17), which will be useful for considering the case of PDMP processes.

Proof. The fact that every measure of finite entropy, relatively to the measure of a jump process, characterizes an other jump process, follows from the Girsanov theorem (see Theorem 2.6 of [51]). Moreover, it is clear that two bursty processes with different drifts and same initial conditions have not the same support, and thus cannot be absolutely continuous one with respect to the other: the relative entropy between two such processes is necessarily infinite.
Thus, for every measure P of finite entropy with respect to $R^{\bar{q}}$, there exists a jump kernel q such that $P \sim R^q$. From the equality (6.16), denoting ξ the solution of the system (6.12) associated to the drift F of the reference process, the corresponding measure of finite entropy with respect to $R^{\bar{q},F}$ corresponds to $\xi_{\#}^{-1} R^q \sim R^{q,F}$. Thus, every measure of finite entropy with respect to $R^{\bar{q},F}$ is the one of a bursty process of the form (6.1).

We now prove the formula (6.17). We first characterize the jump process in terms of martingale. We define for all $t \in [0,1]$:

$$J[a]_t = \sum_{s \leq t, X_s^- \neq X_s^+} a(s, X_s^-, X_s^+).$$

This first lemma comes from classical results about Markov jump processes [22]:

Lemma 16. *Let $R \in \mathcal{P}(càdlàg([0,1], \mathbb{R}^n))$. $R = R^q$ if and only if:*

$$\forall a,\ \exp\left[J[a]_t - \int_0^t \int \{\exp(a(s, X_s, y)) - 1\} q(s, X_s, y) \mathrm{d}y \mathrm{d}s\right] \text{ is a } R\text{-local martingale.}$$

We deduce the following lemma:

Lemma 17. *We define for every $t > 0$ the random variable Z_t by:*

$$Z_t = \exp\left[J\left[\ln\left(\frac{q}{\bar{q}}\right)\right]_t - \int_0^t \{q(s, X_s, \mathbb{R}^n) - \bar{q}(s, X_s, \mathbb{R}^n)\} \mathrm{d}s\right].$$

Then, under the assumption that $\ln\left(\frac{q}{\bar{q}}\right)$ is measurable and bounded on $[0,1] \times \mathbb{R}^{n2}$, Z_t is the Radon-Nikodym derivative of the process of law R^q with respect to the process of law $R^{\bar{q}}$ on $\mathcal{F}_t$.

Proof. Let us denote $P = Z R^{\bar{q}}$. Z is the Radon-Nikodym derivative $\frac{dR^q}{dR^{\bar{q},F_0}}$ if and only if $P = R^q$, which is equivalent from Lemma 16 to:

$$\forall a,\ \exp\left[J[a]_t - \int_0^t \int \{\exp(a(s, X_s, y)) - 1\} q(s, X_s, y) \mathrm{d}y \mathrm{d}s\right] \text{ is a } P\text{-local martingal.}$$

Then we have the equivalence:

$$\begin{aligned}
P = R^q \iff \forall a,\ \exp\Bigg[J\left[a + \ln\left(\frac{q}{\bar{q}}\right)\right]_t - \int_0^t \{q(s, X_s, \mathbb{R}^n) - \bar{q}(s, X_s, \mathbb{R}^n)\} \mathrm{d}s - \\
\int_0^t \int \{\exp\left(a(s, X_s, y)\right) - 1\} q(s, X_s, y) \mathrm{d}y \mathrm{d}s\Bigg] \text{ is a } R^{\bar{q}}\text{-local martingal,} \\
\iff \forall a,\ \exp\Bigg[J\left[a + \ln\left(\frac{q}{\bar{q}}\right)\right]_t - \\
\int_0^t \int \{\exp\left(a(s, X_s, y) + \ln\left(\frac{q}{\bar{q}}\right)\right) - 1\} \bar{q}(s, X_s, y) \mathrm{d}y \mathrm{d}s\Bigg] \text{ is a } R^{\bar{q}}\text{-local martingale.}
\end{aligned}$$

The last equivalence is true thanks to the lemma 16. We have then proved that $P = \frac{dR^q}{dR^{\bar{q},F_0}}$. □

Thus, we obtain:

$$\begin{aligned}
H(R^q|R^{\bar{q}}) &= \mathbb{E}_{R^q}\left[\ln \frac{dR^q}{dR^{\bar{q},F_0}}\right], \\
&= \mathbb{E}_{R^q}\left[J\left[\ln \frac{q}{\bar{q}}\right]_1 - \int_0^1 \{q(s,X_s,\mathbb{R}^n) - \bar{q}(s,X_s,\mathbb{R}^n)\}\mathrm{d}s\right], \\
&= \mathbb{E}_{R^q}\left[\int_0^1 \int \{\frac{q}{\bar{q}}\ln\frac{q}{\bar{q}} + 1 - \frac{q}{\bar{q}}\}\bar{q}(s,X_s,y)\mathrm{d}y\mathrm{d}s\right], \\
&= \mathbb{E}_{R^q}\left[\int_0^1 \int h\left(\frac{q}{\bar{q}}\right)\bar{q}(s,X_s,y)dy\mathrm{d}s\right],
\end{aligned}$$

and the passage from the second line to the third one comes from the fact that $\forall a$:
$J[a]_t - \int_0^t \int a(s,X_s,y)q(s,X_s,y)dy$ is a R^q-local martingale. The formula (6.17) is proved. □

6.2.2 Entropy of the PDMP model

We just sketch the reasoning allowing to deduce the form of the entropy of a PDMP process. A precise analysis is still in construction and we only provide the expected results.
In the following we consider that $a : [0,1] \times {P_E}^2 \times \mathbb{R}^n \to \mathbb{R}$ is a function which is measurable and bounded.

Proposition 18. *Let us consider $R^{Q,F}$ and $R^{\bar{Q},F}$ the laws of two PDMP processes in $[0,1]\times P_E \times \mathbb{R}^n$ related to the jump kernels Q and $\bar{Q}$ and drift F, with the same initial conditions. We have:*

$$H(R^{Q,F}|R^{\bar{Q},F}) = \mathbb{E}_{R^{Q,F}}\left(\int_0^1 \sum_e h\left(\frac{Q}{\bar{Q}}\right)\bar{Q}(t,E_t,e,X_t)\mathrm{d}y\mathrm{d}t\right), \tag{6.18}$$

Sketch of the proof. Once again, we characterize the process $R^{Q,F}$ in terms of local martingale. We define for all $t \in [0,1]$:

$$J[a]_t = \sum_{s\le t, E_t^- \neq E_t^+} a(s,E_t^-,E_t^+,X_t).$$

Wa have the following martingale characterization:

Lemma 19. *Let $R \in \mathcal{P}(càdlàg([0,1],P_E\times\mathbb{R}^n)$. $R = R^{Q,F}$ if and only if:*

$$\forall a,\ \exp\left[J[a]_t - \int_0^t \left(\langle F(s,E_s,E_s),\nabla J[a]_s\rangle + \sum_{e\in P_E}\{\exp(a(s,E_s,e,X_s)) - 1\}Q(s,E_s,e,X_s)dy\right)\mathrm{d}s\right]$$
is a R-local martingale.

Intuitively, the fact that $J[a]_t$ depends only on X_t and not on the whole trajectory $(X_s)_{s\in[0,t]}$ comes from the fact that the knowing $(E_s)_{s\in[0,t]}$, it is possible to reconstruct these trajectories from the value of X_t only. We deduce the following lemma:

Lemma 20. *We define for every $t > 0$ the random variable Z_t by:*

$$\begin{aligned}
Z_t = \exp\Bigg[J\left[\ln\left(\frac{Q}{\bar{Q}}\right)\right]_t - \int_0^t \Big(&\langle F(s,E_s,X_s), J[\ln\left(\frac{Q}{\bar{Q}}\right)]_s(X_s)\rangle \\
&+\{Q(s,E_s,P_E,X_s) - \bar{Q}(s,E_s,P_E,X_s)\}\Big)\mathrm{d}s\Bigg].
\end{aligned}$$

Then, under the assumption that $\ln\left(\frac{Q}{\bar{Q}}\right)$ is measurable and bounded on $[0,1]\times\mathbb{R}^n\times P_E^2$, Z_t is the Radon-Nikodym derivative of the process of law $R^{Q,F}$ with respect to the process of law $R^{\bar{Q},F}$ on $\mathcal{F}_t$.

Thus, using similar reasoning than in the proof of Proposition 14 we obtain the form of (6.18). □

6.2.3 Gamma-convergence of the entropies

In that section, we will slightly abuse the notation and denote $R^{\bar{q},F_0}$ the path measure of a modified bursty process, defining a stochastic process $(0, X_t)_{t\in[0,1]}$, where $(X_t)_{t\in[0,1]}$ is a bursty process. The measure then belongs then the set of path measures $\mathcal{P}(càdlàg([0,1]; P_E \times \mathbb{R}^n))$ for which the random variable on P_E is 0 with probability 1. It is clear that, as this random variable does not bring any information, the formula defining the relative entropy between two such processes is the same as for two bursty processes.

We now prove this corollary of Theorem 11, using the definition of Γ-convergence presented in (8):

Corollary 21. *Let us consider $R_\varepsilon^{\bar{Q},F}$ a measure path characterizing a PDMP process (6.2) as defined in the assumptions of Theorem 11, and $R^{\bar{q},F_0}$ its limit as defined in the same theorem. The functional $F_\varepsilon = H(\cdot|R^{\bar{Q}^\varepsilon,F^\varepsilon}) : \mathcal{P}(càdlàg([0,1]; P_E\times\mathbb{R}^n)) \to \mathbb{R}^+$ Γ-converges to the functional $F = H(\cdot|R^{\bar{q},F_0}) : \mathcal{P}(càdlàg([0,1]; P_E \times \mathbb{R}^n)) \to \mathbb{R}^+$.*

Proof. Under the conditions of Theorem 11, the functional $H(\cdot|R^{\bar{Q}^\varepsilon,F^\varepsilon}) : \mathcal{P}(càdlàg([0,1];\mathbb{R}^n)) \to \mathbb{R}^+$ can be written for every measure path $R_\varepsilon^{Q,F}$ characterizing a PDMP process of finite entropy relatively to $R^{\bar{Q}^\varepsilon,F^\varepsilon}$ and with same initial conditions:

$$H(R^{Q^\varepsilon,F^\varepsilon}|R^{\bar{Q}^\varepsilon,F^\varepsilon}) = \mathbb{E}_{R^{Q^\varepsilon,F^\varepsilon}}\Bigg[\int_0^1 \Big(\sum_{i=1}^n \bar{k}_{on,i} h\left(\frac{k_{on,i}}{\bar{k}_{on,i}}\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=0} + \sum_{i=1}^n \bar{k}_{off,i} h\left(\frac{k_{off,i}}{\bar{k}_{off,i}}\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1}\Big)\mathrm{d}t\Bigg]. \tag{6.19}$$

Denoting R^{q,F_0} the limit of $R^{Q^\varepsilon,F^\varepsilon}$ stated by Theorem 11, we have also:

$$H(R^{q,F_0}|R^{\bar{q},F_0}) = \mathbb{E}_{R^{q,F_0}}\left[\int_0^1 \left(\int_0^\infty \sum_{i=1}^n \bar{k}_{on,i}\bar{P}_u^i\bar{k}_{off,i} h\left(\frac{k_{on,i}P_u^i k_{off,i}}{\bar{k}_{on,i}\bar{P}_u^i\bar{k}_{off,i}}\right)(t, X_t)\mathrm{d}u\right)\mathrm{d}t\right]. \tag{6.20}$$

Step 1: First, we prove that for every measure P in $\mathcal{P}(càdlàg([0,1];\mathbb{R}^n))$ such that $H(P|R^{\bar{q}}) < \infty$, there exists a sequence $(P^\varepsilon)_\varepsilon$ such that $P^\varepsilon \to R^q$ as $\varepsilon \to 0$ for the narrow topology and: $H(P|R^{\bar{q}}) \geq \limsup_{\varepsilon\to 0} H(P|R^{\bar{Q}^\varepsilon,F^\varepsilon})$.

First, we have already seen that for a bursty process, the Girsanov theorem ensures that for every path measure P with finite entropy relatively to $R^{\bar{q},F_0}$, there exists a jump kernel q such that $P \sim R^{q,F_0}$, R^{q,F_0} being the path measure associated to a bursty process characterized by the jump kernel q and the same drift F_0 as $R^{\bar{q},F_0}$ [52]. From Theorem 11, for every kernel q such that $H(R^{q,F_0}|R^{\bar{q},F_0}) < \infty$ we can build a sequence $(R^{Q^\varepsilon,F^\varepsilon})_\varepsilon$ such that $R^{Q^\varepsilon,F^\varepsilon} \to R^{q,F_0}$ for the narrow topology. Thus, we just have to show that for every kernel q such that $H(R^{q,F_0}|R^{\bar{q},F_0}) < \infty$, the sequence $(R^{Q^\varepsilon,F^\varepsilon})_\varepsilon$ is such that: $H(R^{q,F_0}|R^{\bar{q},F_0}) \geq \limsup_{\varepsilon\to 0} H(R^{Q^\varepsilon,F^\varepsilon}|R^{\bar{Q}^\varepsilon,F^\varepsilon})$. We are going to show a stronger result, which is that actually:

$$\lim_{\varepsilon\to 0} H(R^{Q^\varepsilon,F^\varepsilon}|R^{\bar{Q}^\varepsilon,F^\varepsilon}) = H(R^{q,F_0}|R^{\bar{q},F_0}).$$

We consider the test function defined for all gene i and for all $t \in [0,1]$, $(e,x) \in P_E \times \mathbb{R}^n$ by:

$$\begin{cases} \phi^i(t,e,x) := f(t,x) = \int_0^\infty P_u^i k_{off,i} \ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right)(t,x)\mathrm{d}u & \text{if } e_i = 1 \text{ and } \forall j \neq i:\ e_j = 0, \\ \phi^i(t,e,x) = 0 & \text{if not.} \end{cases}$$

Then, we have: Using the martingale characterization associated to the PDMP process, we obtain that for all $0 \geq T \geq 1$,

$$M(T) = f(T, X_T^\varepsilon)\mathbb{1}_{E_{i\,T}^\varepsilon=1} - f(0, X_0^\varepsilon)\mathbb{1}_{E_{i\,0}^\varepsilon=1} - \int_0^T \left(\partial_t f - \frac{F_{1,i}}{\varepsilon}\partial_{x_i} f\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1}\mathrm{d}t - \int_0^T \left(k_{on,i} f(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=0} - \frac{k_{off,i}}{\varepsilon} f \mathbb{1}_{E_{i\,t}^\varepsilon=1}(t, X_t^\varepsilon)\right)\mathrm{d}t$$

is a martingale. From the proof of Theorem 11, we know that E_i^ε converges to 0 weakly, which implies that
$\lim_{\varepsilon\to 0} \mathbb{E}_{R^{Q^\varepsilon,F^\varepsilon}}\left(f(1, X_1^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1} - f(0, X_0^\varepsilon)\mathbb{1}_{E_{i\,0}^\varepsilon=1} - \int_0^T \partial_t f(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1}\mathrm{d}t\right) = 0.$
We then obtain:

$$\lim_{\varepsilon\to 0} \mathbb{E}_{R^{Q^\varepsilon,F^\varepsilon}}\left[\int_0^1 \left(k_{on,i} f(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=0}\right)\mathrm{d}t - \left(\int_0^1 \left(\frac{k_{off,i}}{\varepsilon} f - \frac{F_{1,i}}{\varepsilon}\partial_{x_i} f\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1}\mathrm{d}t\right)\right] = 0.$$

We detail the terms in the second integral for all t, x:

$$\begin{aligned}
\left(k_{off,i} f + F_{1,i}\partial_{x_i} f\right)(t,x) &= \int_0^\infty \left(\left(k_{off,i} P_u^i k_{off,i} - F_{1,i}\partial_{x_i}(P_u^i k_{off,i})\right)\ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right)(t,x)\right)\mathrm{d}u - \\
&\int_0^\infty \left(\left(F_{1,i} P_u^i k_{off,i}\frac{\bar{P}_u^i \bar{k}_{off,i}}{P_u^i k_{off,i}} \times \frac{\bar{P}_u^i \bar{k}_{off,i}\partial_{x_i}(P_u^i k_{off,i}) - P_u^i k_{off,i}\partial_{x_i}(\bar{P}_u^i \bar{k}_{off,i})}{(\bar{P}_u^i \bar{k}_{off,i})^2}\right)(t,x)\right)\mathrm{d}u, \\
&= \int_0^\infty \left(\partial_u(P_u^i k_{off,i})\ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right)(t,x)\right)\mathrm{d}u - \\
&\int_0^\infty \left(\left(P_u^i k_{off,i}\partial_u \ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right) + k_{off,i}(P_u^i k_{off,i}) - \bar{k}_{off,i}(P_u^i k_{off,i})\right)(t,x)\right)\mathrm{d}u, \\
&= \int_0^\infty \partial_u\left((P_u^i k_{off,i})\ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right)(t,x)\right)\mathrm{d}u + \bar{k}_{off,i}(t,x) - k_{off,i}(t,x), \\
&= \bar{k}_{off,i} h\left(\frac{k_{off,i}}{\bar{k}_{off,i}}\right)(t,x).
\end{aligned}$$

where we used the equation 1 of Lemma 12 between the second and the third line. We have then the relation:

$$\lim_{\varepsilon\to 0} \mathbb{E}_{R^{Q^\varepsilon,F^\varepsilon}}\left[\int_0^1 \left(\int_0^\infty k_{on,i} P_u^i k_{off,i}\ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right)(t, X_t^\varepsilon)\mathrm{d}u\mathbb{1}_{E_{i\,t}^\varepsilon=0}\right)\mathrm{d}t - \left(\int_0^1 \bar{k}_{off,i} h\left(\frac{k_{off,i}}{\bar{k}_{off,i}}\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1}\mathrm{d}t\right)\right] = 0.$$

Moreover, it is easy to verify that for all t, x:

$$\bar{k}_{on,i} h\left(\frac{k_{on,i}}{\bar{k}_{on,i}}\right)(t,x) = \bar{k}_{on,i}\bar{P}_u^i \bar{k}_{off,i} h\left(\frac{k_{on,i} P_u^i k_{off,i}}{\bar{k}_{on,i}\bar{P}_u^i \bar{k}_{off,i}}\right)(t,x) - k_{on,i} P_u^i k_{off,i}\ln\left(\frac{P_u^i k_{off,i}}{\bar{P}_u^i \bar{k}_{off,i}}\right)(t,x).$$

Thus, we have for all gene i that:

$$\lim_{\varepsilon\to 0} \mathbb{E}_{R^{Q^\varepsilon,F^\varepsilon}}\left[\int_0^1 \left(\int_0^\infty \bar{k}_{on,i}\bar{P}_u^i \bar{k}_{off,i} h\left(\frac{k_{on,i} P_u^i k_{off,i}}{\bar{k}_{on,i}\bar{P}_u^i \bar{k}_{off,i}}\right)(t, X_t^\varepsilon)\mathrm{d}u\mathbb{1}_{E_{i\,t}^\varepsilon=0}\right)\mathrm{d}t - \left(\int_0^1 \bar{k}_{on,i} h\left(\frac{k_{on,i}}{\bar{k}_{on,i}}\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=0}\mathrm{d}t + \int_0^1 \bar{k}_{off,i} h\left(\frac{k_{off,i}}{\bar{k}_{off,i}}\right)(t, X_t^\varepsilon)\mathbb{1}_{E_{i\,t}^\varepsilon=1}\mathrm{d}t\right)\right] = 0.$$

We finally deduce, using the formula (6.19), that:

$$\lim_{\varepsilon\to 0} H(R^{Q^\varepsilon,F^\varepsilon}|R^{\bar{Q}^\varepsilon,F^\varepsilon}) = \lim_{\varepsilon\to 0} \mathbb{E}_{R^{Q^\varepsilon,F^\varepsilon}}\left[\int_0^1 \left(\int_0^\infty \sum_{i=1}^n \bar{k}_{on,i}\bar{P}^i_u\bar{k}_{off,i} h\left(\frac{k_{on,i}P^i_u k_{off,i}}{\bar{k}_{on,i}\bar{P}^i_u\bar{k}_{off,i}}\right)(t,X^\varepsilon_t)du\right)\mathbb{1}_{E^\varepsilon_{i\,t}=0}\mathrm{d}t\right].$$

We remark that the quantity inside the integral on the right-hand side is exactly the quantity appearing in the formula (6.20). Then, the convergence in law of $(E^\varepsilon_t, X^\varepsilon_t)_t$ to $(0, X^\varepsilon_t)$ allows to conclude.

Step 2: Second, we prove that for every sequence $(P^\varepsilon)_\varepsilon$ of measures in $\mathcal{P}(càdlàg([0,1];P_E\times\mathbb{R}^n))$ such that $P^\varepsilon \to P$ narrowly, we have $H(P|R^{\bar{q},F_0}) \leq \liminf_{\varepsilon\to 0} H(P^\varepsilon|R^{\bar{Q}^\varepsilon,F^\varepsilon})$.

We are going to use the dual formulation of the entropy, that we presented in Chapter 5. We recall first that the Legendre transform of $h(u) = u\log u$ is $h^*(v) = e^{v-1}$, and that we have in particular for all u, v:

$$u\log u \geq uv - e^{v-1}.$$

Then, for function $G \in C_b(càdlàg([0,1];P_E\times\mathbb{R}^n))$, taking $u = \frac{\mathrm{d}P^\varepsilon}{\mathrm{d}R^{Q^\varepsilon,F^\varepsilon}}$ and $v = G$, integrating with respect to $R^{Q^\varepsilon,F^\varepsilon}$ we obtain:

$$\begin{aligned} H(P^\varepsilon|R^{\bar{Q}^\varepsilon,F^\varepsilon}) &= \int \frac{\mathrm{d}P^\varepsilon}{\mathrm{d}R^{Q^\varepsilon,F^\varepsilon}}\log\frac{dP^\varepsilon}{\mathrm{d}R^{Q^\varepsilon,F^\varepsilon}}\mathrm{d}R^{Q^\varepsilon,F^\varepsilon}, \\ &\geq \int \frac{dP^\varepsilon}{\mathrm{d}R^{Q^\varepsilon,F^\varepsilon}} G\mathrm{d}R^{Q^\varepsilon,F^\varepsilon} - \int e^{G-1}\mathrm{R}^{\bar{Q}^\varepsilon,F^\varepsilon}, \\ &= \int G\mathrm{d}P^\varepsilon - \int e^{G-1}\mathrm{R}^{\bar{Q}^\varepsilon,F^\varepsilon}. \end{aligned}$$

Moreover, as $P^\varepsilon \to P$ and $R^{\bar{Q}^\varepsilon,F^\varepsilon} \to R^{q,F_0}$ narrowly (by Theorem 6.1 for the second one), we have for all $G \in C_b(càdlàg([0,1];P_E\times\mathbb{R}^n))$:

$$\int G\mathrm{d}P^\varepsilon \to \int G\mathrm{d}P, \quad \int e^{G-1}\mathrm{R}^{\bar{Q}^\varepsilon,F^\varepsilon} \to \int e^{G-1}\mathrm{d}R^{\bar{q},F_0}.$$

Thus, we obtain:

$$\liminf_{\varepsilon\to 0} H(P^\varepsilon|R^{Q^\varepsilon,F^\varepsilon}) \geq \int G\mathrm{d}P \int e^{G-1}\mathrm{d}R^{\bar{q},F_0},$$

and we can conclude passing to the sup on the right-hand side.

□

In conclusion, minimizing the entropy relatively to the bursty model is equivalent to minimizing the entropy relatively to the PDMP model in the limit $\varepsilon \to 0$.

6.3 Dual of the Schrödinger problem when the reference is a bursty process

In this section, we build the dual formulation of the dynamical Schrödinger problem when the reference is a bursty process. It appears similar to the dual of the Schrödinger problem when the reference is a coupling associated to the bursty process. The relation between the Schrödinger potentials and the jump kernel characterizing the solution of the dynamical Schrödinger problem will provide a method for building this kernel from two temporal observations.

Remark that in this section, most of the analyses are formal: the results that are presented will nevertheless allow to motivate some numerical applications afterwards.

6.3.1 Dual of the dynamical Schrödinger problem

We now establish the dual formulation of the dynamical Schrödinger problem given two marginal distributions μ and ν at times 0 and 1. For a reference kernel $\bar{q}$, a drift F, we recall that any process of finite entropy with respect to a bursty process $R^{\bar{q},F}$ is a bursty process itself, which is then characterized by a jump kernel q, the same drift F, and its distribution ρ which verifies the master equation:

$$\forall t, x : \partial_t \rho(t,x) = div(F\rho)(t,x) + \int \rho(t,z) q(t,dz,x) - \rho(t,x) \int q(t,x,dz).$$

From the Proposition 6.17, we have then the following formulation for the Schrödinger problem relatively to $R^{\bar{q},F}$ on $[0,1]$, with two probability measures μ and ν at times 0 and 1 respectively:

$$\begin{aligned}\mathrm{Sch}(R^{\bar{q},F};\mu,\nu) = \inf_{\rho,q} \Big\{ \int_0^1 \int\int h\left(\frac{q}{\bar{q}}\right) \bar{q}(t,x,\mathrm{d}y)\rho(t,\mathrm{d}x)\mathrm{d}t \,\big|\, \\ \forall t, x : \partial_t \rho(t,x) = div(F\rho)(t,x) + \int \rho(t,z) q(t,dz,x) - \rho(t,x) \int q(t,x,dz), \\ \rho(0,\cdot) = \mu(\cdot),\ \rho(1,\cdot) = \nu(\cdot) \Big\}\end{aligned}$$

Proposition 22. *For a reference kernel $\bar{q}$, a drift F, and two probability measures μ and ν, the dual of the dynamical Schrödinger problem on $[0,1]$ can be written:*

$$\begin{aligned}\mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu) := \sup_{\phi(t,\cdot)\in C_c^\infty} \{ \int \phi(1,y)\nu(\mathrm{d}y) - \int \phi(0,x)\mu(\mathrm{d}x) \ \big| \\ \partial_t e^{\phi(t,x)} + \langle F(t,x), \nabla e^{\phi(t,x)} \rangle = \int \left(e^{\phi(t,x)} - e^{\phi(t,z)} \right) \bar{q}(t,x,dz) \}.\end{aligned} \tag{6.21}$$

Moreover, for the optimum q^* and ϕ^* of both problems, we necessarily have for all $t > 0$, $x, z \in \mathbb{R}^n$ the relation:

$$\frac{q^*}{\bar{q}}(t,x,z) = e^{\phi^*(t,z) - \phi^*(t,x)}. \tag{6.22}$$

Remark 23. We realized that the form of this dual formulation is the same of the general form proved in Gentil et al. [32] for Markov processes. Indeed, the PDE defining e^ϕ in (6.21) is the backward equation associated to a bursty process with kernel $\bar{q}$ and drift F: for all t, x we have

$$\partial_t e^{\phi(t,x)} = -\mathcal{A}_{\bar{q},F} e^{\phi(t,x)}.$$

Then, if we denote ϕ^* the solution of the dual dynamical problem (6.21), we obtain that for all $x \in \mathbb{R}^n$:

$$\phi^*(0,\cdot) = \ln T_1^{\bar{q},F} e^{\phi^*(1)}(\cdot), \tag{6.23}$$

where $(T_t^{\bar{q},F})_{t\in[0,1]}$ is the semigroup associated to the reference bursty process, which allow to retrieve the form presented in Gentil et al.[78]. We nevertheless derive the non trivial relation (6.22), which will appear in the dual construction.

Proof. We denote ϕ a Lagrange multiplier such that for all $t \in [0,1]$, $\phi(t,\cdot) \in C_c^\infty(\mathbb{R}^n)$. We replace the condition on the master equation linking ρ and q on the formula of $\mathrm{Sch}(R^{\bar{q},F};\mu,\nu)$ by a term of the form:

$$\sup_\phi \int_0^1 \left(\partial_t \rho(t,x) = div(F\rho)(t,x) + \int \rho(t,z) q(t,dz,x) - \rho(t,x) \int q(t,x,dz) \right) \phi(t,x) dt.$$

In plain word, this forces the master equation to be verified for almost all t, x unless the the value of $\mathrm{Sch}(R^{\bar{q},F};\mu,\nu)$ is infinite. Integrating by parts w.r.t t, we obtain:

$$\begin{aligned}
\mathrm{Sch}(R^{\bar{q},F};\mu,\nu) = \inf_{\rho,q}\Bigg\{ & \int_0^1 \int\int h\left(\frac{q}{\bar{q}}\right)\bar{q}(t,x,\mathrm{d}y)\rho(t,\mathrm{d}x)\mathrm{d}t + \sup_{\phi}\Big\{ \int \phi(1,y)\rho(1,\mathrm{d}y) - \int \phi(0,x)\rho(0,dx) \\
& - \int_0^1 \int \big[\partial_t\phi(t,x) + \langle F(t,x), \nabla\phi(t,x)\rangle - \int (\phi(t,z)-\phi(t,x))q(t,x,dz)\big]\rho(t,dx)\mathrm{d}t\Big\} \ | \\
& \rho(0,\cdot)=\mu(\cdot), \rho(1,\cdot)=\nu(\cdot)\Bigg\}, \\
\geq \sup_{\phi}\Bigg\{ & \int \phi(1,y)\rho(1,\mathrm{d}y) - \int \phi(0,x)\rho(0,\mathrm{d}x) + \inf_{\rho,q}\Big\{ \int_0^1 \int \Big[\int h(\frac{q}{\bar{q}})\,\bar{q}(t,x,dz) \\
& - \partial_t\phi(t,x) - \langle F(t,x), \nabla\phi(t,x)\rangle - \int (\phi(t,z)-\phi(t,x))q(t,x,dz)\big]\rho(t,dx)\mathrm{d}t \ | \\
& \rho(0,\cdot)=\mu(\cdot), \rho(1,\cdot)=\nu(\cdot)\Big\}\Bigg\} := \mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu).
\end{aligned}$$

The sign $\geq$ comes from the fact that an inf sup is always greater than a sup inf. Remark that an equality would hold at the second line provided that we could invert the sup and the inf at the second line, but we have not yet explored this direction.

Then, by denoting $f(t,x,z) = \frac{q(t,x,z)}{\bar{q}(t,x,z)}$, we obtain:

$$\begin{aligned}
\mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu) = \sup_{\phi}\Bigg\{ & \int \phi(1,y)\rho(1,\mathrm{d}y) - \int \phi(0,x)\rho(0,dx) + \\
& \inf_{\rho,f}\Big\{ \int_0^1 \int \Big[- \partial_t\phi(t,x) - \langle F(t,x), \nabla\phi(t,x)\rangle + \int \big[(f\ln f + 1 - f)(t,x,z) - \\
& f(t,x,z)\times(\phi(t,z)-\phi(t,x))\big]\bar{q}(t,x,dz)\Big]\rho(t,dx)\mathrm{d}t \ | \\
& \rho(0,\cdot)=\mu(\cdot), \rho(1,\cdot)=\nu(\cdot)\Big\}\Bigg\}.
\end{aligned}$$

Deriving with respect to f, we obtain at the optimum the following relation:

$$\ln f(t,x,z) = \phi(t,z) - \phi(t,x),$$

from which the optimality condition (6.22) follows.
Replacing into the equation, we finally obtain:

$$\begin{aligned}
\mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu) = \sup_{\phi}\Bigg\{ & \int \phi(1,y)\rho(1,\mathrm{d}y) - \int \phi(0,x)\rho(0,dx) - \\
& \inf_{\rho}\Big\{ \int_0^1 \int \big[\partial_t\phi(t,x) + \langle F(t,x), \nabla\phi(t,x)\rangle - \\
& \int \left(1 - e^{\phi(t,z)-\phi(t,x)}\right)\bar{q}(t,x,dz)\big]\rho(t,dx)dy \ | \ \rho(0,\cdot)=\mu(\cdot), \rho(1,\cdot)=\nu(\cdot)\Big\}\Bigg\}, \\
= \sup_{\phi}\Bigg\{ & \int \phi(1,y)\nu(\mathrm{d}y) - \int \phi(0,x)\mu(\mathrm{d}x) \ | \\
& \partial_t\phi(t,x) + \langle F(t,x), \nabla\phi(t,x)\rangle = e^{-\phi(t,x)}\int \left(e^{\phi(t,x)} - e^{\phi(t,z)}\right)\bar{q}(t,x,dz)\Bigg\},
\end{aligned}$$

where the passage from the first to the second equality comes from the fact that the inf on ρ is necessarily infinite if ϕ is not solution of the PDE at the last line. We finally deduce the dual formulation (6.21). □

6.3.2 Formal relation with the non-dynamical Schrödinger problem

As mentioned in Section 1.1.2, there is an equivalence between the "non-dynamical" Schrödinger problem (1.4) and the dynamical one (1.5) in that meaning that they are associated to the same objective values and that we can link the solutions of the two problems by the formula (1.6). However, it is not clear at the moment how to find the function ϕ characterizing the optimal jump kernel in (6.22). Denoting R_{01}^* the solution of (1.4) when the reference kernel is R_{01}, we are going to show, at least formally, that the two entropic potentials characterizing the Radon-Nikodim derivative $\frac{dR_{01}^*}{dR_{01}}$ allow to define this kernel.

First, we recall some results about the dual of the non-dynamical Schrödinger problem (1.4) that can be found in [70] (and the references herein). We begin by presenting the dual formulation of the Schrödinger problem:

Proposition 24. *We have:*

$$\inf_{P\in\Gamma(\mu,\nu)} H(P|R_{01}) = \sup_{\phi\in L^1(\mu),\psi\in L^1(\nu)} \int \phi(x)\mu(\mathrm{d}x) + \int \psi(y)\nu(\mathrm{d}y) + 1 - \int e^{\phi(x)+\psi(y)} R_{01}(x,y)dxdy.$$

We are going to assume that the solution R_{01}^* has a density of the form

$$\frac{\mathrm{d}R_{01}^*}{\mathrm{d}R_{01}} = e^{\phi^*\oplus\psi^*},\ R_{01} - as, \tag{6.24}$$

where $\phi^* \in L^1(\mu)$ and $\psi^* \in L^1(\nu)$ are called the Schrödinger potentials, we have the following corollary:

Corollary 25. *Assuming the relation (6.24), the Schrödinger potentials phi* and ψ^* are the maximizers of the dual problem and we have at the optimum:¡*

$$H(R_{01}^*|R_{01}) = \int \phi^*(x)\mu(\mathrm{d}x) + \int \psi^*(y)\nu(\mathrm{d}y).$$

Moreover, it is also known that, still assuming (6.24), e^{ϕ^*} and e^{ψ^*} are the limits of the Sinkhorn algorithm, that solve the system:

$$\begin{cases} \forall x,\ e^{\phi^*(x)} &= \frac{\mu(x)}{\int e^{\psi^*(y)}R_{01}(x,y)dy}, \\ \forall y,\ e^{\psi^*(y)} &= \frac{\nu(y)}{\int e^{\phi^*(x)}R_{01}(x,y)dx}, \end{cases} \tag{6.25}$$

We now develop a formal reasoning by associating, for all $t \in [0,T]$ and $x \in \mathbb{R}^n$, the transition probability kernel associated to the reference measure $R^{\bar{q},F}$ at time t, $p_t^{\bar{q},F}(x,\mathrm{d}y)$, which is a probability measure on $\mathbb{R}^n$, to its probability density function $p_t^{\bar{q},F}(x,y)$. We aim to show, formally, that the solutions ϕ^* and ψ^* of the system (6.25) for $R_{01} = p_1^{\bar{q},F}\mu$, characterizing the solution of $\mathrm{Sch}(p_1^{\bar{q},F}(x,y)\mu(x);\mu,\nu)$, allow to build a function ϕ solution of the dual's dynamical problem $\mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu)$. We consider the function $\phi : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ such that:

$$\forall t \in [0,1] :\ \phi(t,\cdot) = \ln T_{1-t}^{\bar{q},F} e^{\psi^*}(\cdot). \tag{6.26}$$

where we recall that $(T_t^{\bar{q},F})_{t\in[0,1]}$ is the semigroup associated to $R^{\bar{q},F}$.
We observe first that the system (6.25) becomes:

$$\begin{cases} \forall x,\ e^{\phi^*(x)} &= \frac{1}{T_1^{\bar{q},F}e^{\psi^*(x)}}, \\ \forall y,\ e^{\psi^*(y)} &= \frac{\nu(y)}{\int e^{\phi^*(x)}p_1^{\bar{q},F}(x,y)\mu(\mathrm{d}x)}, \end{cases} \tag{6.27}$$

Thus, the first line implies that we have $\phi(0,\cdot) = -\phi^* = \ln T_1^{\bar{q},F} e^{\phi^*(1,x)}$, and then the condition (6.23) between $\phi(0)$ and $\phi(1)$ is well verified with $\phi(0,\cdot) = -\phi^*$ and $\phi(1,\cdot) = \psi^*$.

Second, the objective values of the non-dynamical and dynamical problems being the same, we necessarily have by Corollary 25:

$$\mathrm{Sch}(R^{\bar{q},F};\mu,\nu) = \int \phi^*(x)\mu(\mathrm{d}x) + \int \psi^*(y)\nu(\mathrm{d}y).$$

As $\mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu) \leq \mathrm{Sch}(R^{\bar{q},F};\mu,\nu)$, and that we have seen that the function ϕ defined in (6.26) with $\phi(0) = -\phi^*$ and $\phi(1) = \psi^*$ is compatible with the constraints of the dual problem, we have necessarily:

$$\mathrm{Sch}^*(R^{\bar{q},F};\mu,\nu) = \int \phi^*(x)\mu(\mathrm{d}x) + \int \psi^*(y)\nu(\mathrm{d}y).$$

We also remark that, still formally, the second line of the system (6.27) implies that defining

$$p_1^{q^*,F}(x,y) = e^{\phi(1,y)-\phi(0,x)} p_1^{\bar{q},F}(x,y),$$

which corresponds to the condition (6.22) for the optimal jump kernel in the case of a bursty process, the resulting distribution ρ^* verifies well $\rho^*(1,\cdot) = \nu$ when $\rho^*(0,\cdot) = \rho(0,\cdot) = \mu$. Indeed we have:

$$\begin{aligned}
\rho^*(1,\mathrm{d}y) &= \int p_1^{q^*}(x,\mathrm{d}y)\mu(\mathrm{d}x),\\
&= \int e^{\phi^*(x)} e^{\psi^*(y)} p_1^{\bar{q}}(x,\mathrm{d}y)\mu(\mathrm{d}x),\\
&= \int e^{\phi^*(x)} \frac{\nu(\mathrm{d}y)}{\int e^{\phi^*(x)} p_1^{\bar{q},F}(x,\mathrm{d}y)\mu(\mathrm{d}x)} p_1^{\bar{q},F}(x,\mathrm{d}y)\mu(\mathrm{d}x),\\
&= \nu(\mathrm{d}y).
\end{aligned}$$

Remark 26. We emphasize that this formal analysis does not take into account difficulties that may appear, in particular when considering the second line of (6.27). Indeed, $p_t(x,\mathrm{d}y)$ on the right-hand side is actually a measure, although on the left-hand side we have a function. The most common hypothesis used for tackle this problem is that the underlying stochastic process is revertible, like for the revertible Brownian motion, which is not the case for the bursty process.

6.3.3 Partial conclusion

We recall that we developed in Chapter 5 a numerical method, combining a preprocessing step and the Sinkhorn algorithm, in order to find from from any triplet $(R_{01};\mu,\nu)$ a coupling $R_{01}^* \in \Pi(\mu^*,\nu^*)$, with μ^* and ν^* defined by the formula (12) of Chapter 5, and a, b two strictly non-negative functions such that:

$$\forall x, y,\ R^*(x,y) = a(x)b(y)R(x,y),$$

as soon as $R^*(x,y) > 0$.
In that case, the Schrödinger potentials ϕ^* and ψ^* of $\mathrm{Sch}(R,\mu^*,\nu^*)$ can then be approximated by $\ln a$ and $\ln b$ on this support. Indeed, they correspond to the potentials of the Schrödinger problem $\mathrm{Sch}(R_{01};\mu^*,\nu^*)$, which is the problem having a solution whose marginals are the closest to μ and ν.
Identifying $\phi(0)$ to ϕ^* and $\phi(1)$ to ψ^* in (6.21), and taking $R(x,y)\mathrm{d}x\mathrm{d}y = p_t(x,\mathrm{d}y)\mu(\mathrm{d}x)$, we can then deduce a method to compute the optimal jump kernel q^* associated to the solution of the problem $\mathrm{Sch}(R^{\bar{q},F};\mu^*,\nu^*)$:

1. Find the Schrödinger potential ϕ^* and ψ^* associated to the Schrödinger problem $\mathrm{Sch}(p_t(x,\mathrm{d}y)\mu(\mathrm{d}x);\mu^*,\nu^*)$ on the support of R^*;

2. Set $\forall x,y,t \in \mathbb{R}^n \times \mathbb{R}^n \times [0,1]$, $q^*(t,x,y) = \frac{T^{\bar{q},F}_{1-t}b(y)}{T^{\bar{q},F}_{1-t}b(x)}\bar{q}(t,x,y)$.

In particular, we have $q^*(1,x,y) = \frac{b_1(y)}{b_1(x)}\bar{q}(1,x,y)$.

We obtain a formula allowing to compute the time-dependent jump kernel characterizing the optimal process, solution of the dynamical Schrödinger problem, at the time of the second observation. According to the interpretation of the entropy regarding the Sanov theorem, developed in Section 1.1.2, this formula allows thus to characterize the optimal modifications that are needed for the process to be compatible with the observations.

However, two difficulties appear for answering the questions that we addressed in the introduction of Part III:

- The optimal kernel q^* is not parametric, and then does not allow to be directly linked to the mechanistic parameters of the model (1.17), and in particular an underlying optimal GRN for inference purposes;

- There are some numerical difficulties related to the fact that the bursty model has no close formula for characterizing the semigroup $T^{\bar{q}}_t$ and its transition probability kernel.

These points will be addressed in the next section. We will then show preliminary applications on *in silico* generated datasets.

6.4 Practical aspect

We now consider that we are observing two sets of independent cells S_0 and S_T at $t = 0$ and $t = T$ (in hours). We also consider that we have a prior knowledge on the system, under the form of a set of parameters calibrating the PDMP process. In particular, we denote θ^R and k^R_{off} the matrix describing the GRN calibrating the functions $k_{on,i}$, and the vector describing the switching rates from the state *off* to the state *on* for every gene i. The aim of this section is to describe an numerical method for estimating the optimal network θ^* and rate k^*_{off} w.r.t the observations and the reference parameters.
Relying on the results about the Schrödinger problem previously shown, we propose a 3-steps method which consists in:

- Solving the Schrödinger problem on appropriate starting and ending spaces, with a bursty reference process associated to a Kernel $q(x, x + he_i) = k^{\theta^R}_{on,i}(x)\frac{k_{off,i}}{s_i}e^{-\frac{k_{off,i}}{s_i}h}$;

- Finding the optimal kernel q^*, and using the form (6.5) for the optimal jump kernel, the associated optimal burst rate functions $k^*_{on,i}$ and $k^*_{off,i}$;

- Finding the optimal associated parameters at time T, and in particular θ^*, of the associated PDMP process.

The second step requires some assumptions on the dependence in time and space of the burst rate functions, in order to be able to identify two sets of functions $k_{on,i}$ and $k_{off,i}$ from the formula (6.5) which is a combination of them. The third step consists in considering a parametric model for these functions, which could be then inferred by a specific set of regression problems as we did in Chapter 3.

Remark 27. As we only consider two timepoints, we would like to consider a dataset generated by a bursty model such that the parameters are likely to be recovered from the observation of these two timepoints. That is the case of a toggle-switch network (two genes which activate themselves and inhibit each other), similar to the one used throughout Chapter 2 and described in Appendix C of this chapter. Indeed, we have shown that CARDAMOM was able to infer accurately such GRN from the observations of two timepoints (see Chapter 3, Figure 6). We will thus consider datasets generated using the bursty model (1.17) and calibrated by this toggle-switch GRN in the following.

6.4.1 Starting and ending spaces of interest

In order to link this framework to the analyses of Section 6.3, the sets of cells S_0 and S_T are described by two probability distributions ρ_0 and ρ_T. For the set S_0, it is natural to use the empiric distribution, *i.e* to consider:

$$\rho_0(\mathrm{d}x) = \frac{1}{N_0} \sum_{X \in S_0} \delta_X(\mathrm{d}x),$$

where $N_0 = |S_0|$.
It is however not sufficient to consider a discrete ending space. Indeed, as it will be justified more precisely afterwards in Proposition 28, we need for estimating the optimal jump kernel, at time $t = T$, a minimum amount of vectors of $\mathbb{R}^n$ that verify together a certain condition, that can be not found in S_T. We thus need to extend the space. For this sake, we convolve the distribution $\hat{\rho}_T(\mathrm{d}y) = \sum_{Y \in S_T} \delta_Y(\mathrm{d}y)$ against a Gaussian of width h. We denote ρ_T^h the resulting distribution. We then consider a discretization of the gene expression space, and build a gird denoted G_N^n ,of size N^n, where N corresponds to the number of bins in each direction, and n the number of genes. Every point on this gird then corresponds to a vector of $\mathbb{R}^n$, where each coordinate i has its value in $\{y_0^i, \cdots, y_N^i\}$. For example, we may choose:

$$y_0^i = \min_{Y \in S_T} Y_i,\ y_N^i = \max_{Y \in S_T} Y_i, \text{ and } \forall 1 \leq k \leq N :\ y_k^i = y_0^i + \frac{k}{N}(y_N^i - y_0^i).$$

Finally with the notations of the Schrödinger problem, we obtain that:

- μ is a vector of size N_0 for which each entry is equal to $\frac{1}{N_0}$;
- ν is a vector of size N^n, and for all y in G_N^n: $\nu(y) = \rho_T^h(y)$.

6.4.2 Approximating the Kernel associated to the bursty process

A major numerical difficulty when studying the Schrödinger problem with the bursty process in $[0, T]$ as a reference, is that the reference measure between two timepoints, R_{0T}, has no explicit analytical form. We recall that the quantity of interest for building R_{0T} is $p_T(x, \mathrm{d}y)$, where p_T is the transition probability kernel associated to the bursty process at time T. Thus, before solving the Schrödinger problem, we must be able to estimate this coupling. For this, we used two different methods. On one hand, we built a numerical scheme for solving the master equation of the bursty process. Starting from any cell $x \in S_0$ at time 0, we then estimated the exact value of $p_T(x, \mathrm{d}y)$ on the grid G_N^n, allowing to build the vector ν on this grid. However, solving this numerical scheme may be computationally very intensive when the number of genes is high. An alternative method consisted in estimating $p_T(x, \mathrm{d}y)$ using a Monte-Carlo method: we simulated from any cell $x \in S_0$ at time 0 a certain number of realization of the process. Convolving the empirical distribution obtained $\hat{\rho}_T(dy|x)$ against a Gaussian of width h, we then obtained an estimation of $p_T(x, \mathrm{d}y)$ on the grid G_N^n.

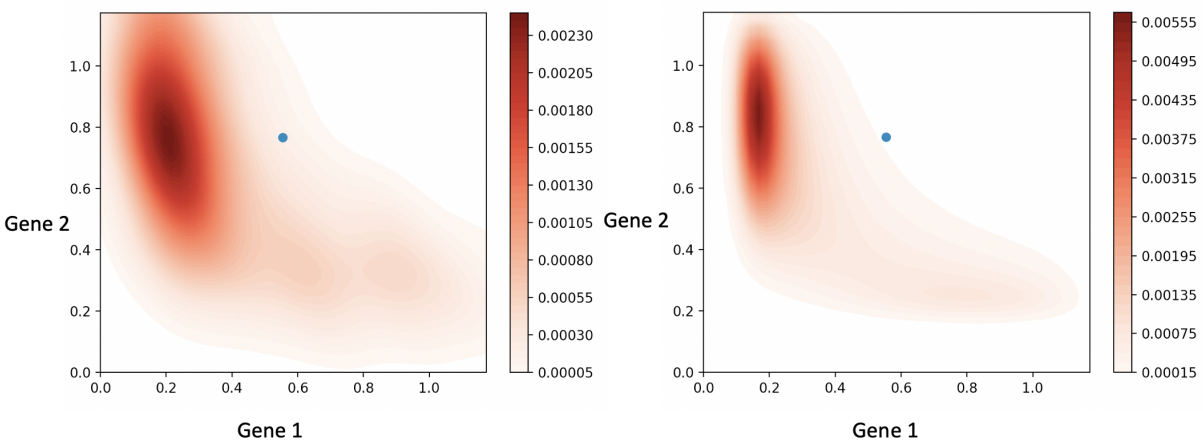


Figure 6.1: Comparison of the kernel associated to the mechanistic model from a cell (the blue dot in the figure), estimated from the exact resolution of the master equation (on the right) and a Monte-Carlo method with 200 simulated cells (on the left).

We compare in Figure 6.1 the transition probability kernel $p_T(x, \mathrm{d}y)$ obtained with the two methods, from a certain x in $[0,1]^2$ (represented by a blue dot). As mentioned at the beginning of this section, we used as a reference the bursty process calibrated by a toggle-switch network. In the following, we will choose the second method for the numerical applications, which is easier and faster to use than the first one when considering many genes.

6.4.3 Solving the Schrödinger problem

With the reference coupling $R_{0T} = p_T(x, dy)\mu(\mathrm{d}x)$ estimated for all $x \in S_0$, we can now solve the Schrödinger problem $\mathrm{Sch}(R_{0T}; \mu, \nu)$, with μ, ν defined as previously. The results of Chapter 5 ensure that the Sinkhorn algorithm will converge, and that taking the geometric mean of the two limits defined in Theorem 12 of Chapter 5, we are able to recover an optimal coupling with respect to the data. We represented in Figure 6.2 the evolution of the minimal relative entropy $\min_{R \in \Pi(\mu,\nu)} H(R|R_{0T})$ in two cases: when the transition probability kernel $p_T(x, dy)$ is the one of a bursty process with a null network (*i.e* the functions $k_{on,i}$ are scalar), in blue, and when it is the one of the right network, *i.e* the one used for simulating the datasets S_0 and S_T, in red. We observe that from $T = 2h$, the metric is able to distinguish the wrong reference process from the right one: the minimal relative entropy, which can be interpreted as a kind of distance between the reference process and the observations, increases almost linearly with respect to the time gap between the observations in the first case, while it remains close to 0 in the second case, even for a large time gap.

6.4.4 Computing an optimal kernel and inferring an associated gene regulatory network

We now aim to deduce, from the optimal kernel measured at a time $t = T$, an estimation of the optimal burst rate functions $k^*_{on,i}$ at this time. For this, we need to make assumptions on the kernel associated to the optimal bursty process. From the convergence result of Theorem 11, we can state the following simplification of Proposition 22:

Proposition 28. *If the burst rate functions $k^*_{off,i}$ calibrating the optimal process with respect to the data are constant in time and space, the associated optimal functions $k^*_{on,i}$ verify the relation*

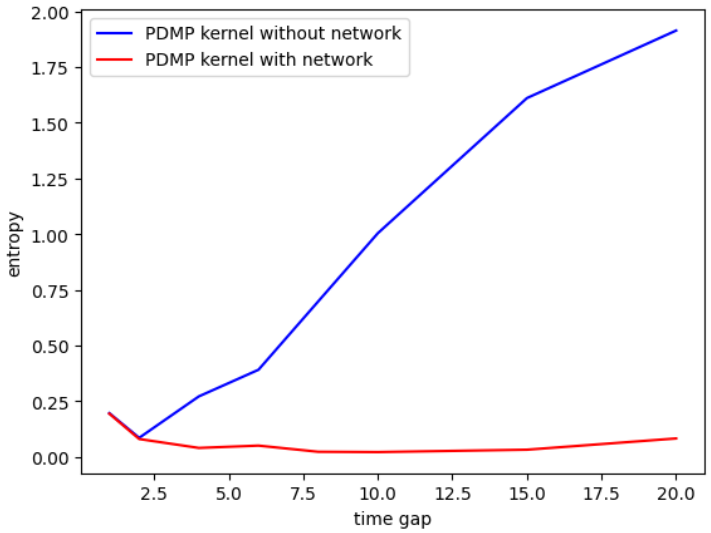


Figure 6.2: Evolution of the minimal entropy associated to the optimal coupling as a function of the time gap between the two timepoints, when the reference kernel is computed with the right network (in red) and a null network (in blue).

at every T*:*

$$\forall (x,y) \in \mathbb{R}^n : \; k^{\theta^*}_{on,i}(T,x) = \frac{b_T(y)}{b_T(x)} \times \underbrace{k^{\theta}_{on,i}(x)}_{\text{reference GRN}} \times \underbrace{\frac{k_{off,i}}{k^*_{off,i}} e^{\frac{k^*_{off,i}-k_{off,i}}{s}h}}_{\text{scaling of mRNA bursts}}, \tag{6.28}$$

if $\exists h > 0$ *such that* $\forall j \neq i, x_j = y_j$*, and* $y_i = x_i + h$*.*

Remark 29. This proposition is the direct consequence of Proposition 22, under the hypothesis that the rates $k^*_{off,i}$ does not depend on space. However, it is easy to see that the value of y does not play any role and should be always compensated by $b_T(Y)$. We should have for all $x, h > 0$ and every gene i:

$$\frac{\mathrm{d}}{\mathrm{d}h}\left[b_T(x + he_i) e^{\frac{k^*_{off,i}-k_{off,i}}{s_i}h} \right] = 0, \tag{6.29}$$

where e_i is a vector with all coordinates are equal to 0 except the i^{th} which is equal to 1. Although we are going to work under this assumption, we have to keep in mind that all the results that we deduced from Proposition (28) may be inaccurate if the relation (6.29) is not verified.

In particular, from the grid G^n_N, we could compute en error for each gene i:

$$err_i = \sum_{l=1}^{N} \left[\sum_{k_1, y^i_{k_1} > y^i_l} \sum_{k_2, y^i_{k_2} > y^i_l} | \, b_T(y^i_{k_1}) e^{\frac{k^*_{off,i}-k_{off,i}}{s_i}(y^i_{k_1} - y^i_l)} - b_T(y^i_{k_2}) e^{\frac{k^*_{off,i}-k_{off,i}}{s_i}(y^i_{k_2} - y^i_l)} \, | \right],$$

which would quantify how wrong the hypothesis that $k^*_{off,i}$ is a scalar.

From Proposition 28, we can deduce an approximate method for estimating the optimal network θ and rates $k_{off,i}$ at $t = Th$ associated to two sets of data S_0 and S_T, and a reference PDMP process associated to a network θ^R and burst rates $k^{\theta^R}_{on}, k^R_{off}$. Considering that the burst rate functions $k^*_{off,i}$ calibrating the optimal process with respect to the data are constant in time and space, it consists in solving the following minimization problem:

$$\theta^*, k^*_{off} = \underset{\theta, k_{off}}{\arg\min} \sum_{x \in G^n_N} \sum_{i=1}^{n} F_i(\theta, k_{off,i}, x, b) + \lambda |\theta - \theta^R|, \tag{6.30}$$

where λ is a penalization coefficient and for all gene i:

$$F_i(\theta, k_{off,i}, x, b) = \sum_{y \in G_N^n} \left(b_T(x) k_{on,i}^{\theta}(x) k_{off,i} - b_T(y) k_{on,i}^{\theta^R}(x) k_{off,i}^R e^{\frac{k_{off,i} - k_{off,i}^R}{s} h} \right)^2 \mathbb{1}_{y = x + h e_i}.$$

Solving these problems then allows to find the optimal θ^* and $k_{off,i}^*$ regarding the data.

We represent in Figure 6.3b the results obtained by this method for reconstructing the optimal θ^*, depending on the time gap between the two timepoints, when the right network is still a toggle-switch network of two genes. We observe that the network obtained when the reference process is a bursty process calibrated by a null network becomes closer to the right network when the time gap increases: this is due to the fact that the process needs time to equilibrate, *i.e* for final measure to be far enough from the first one to be representative of the effect of the network. However, the distance to the right network (in total variation) remains not so small even after a certain time. More precisely, we observe that the method is able to detect the inhibitions of the network, but not the activations, or at least not with the right intensity.
When the reference process is calibrated with the right network (the one used for generating the data), we observe that except for too small time gaps (when the data is not representative of the dynamics induced by the network), the network inferred is close to the right one. This was expected from the results of Figure 6.2: the relative entropy between the optimal coupling and the reference one was close to 0 for any value of the time gap.

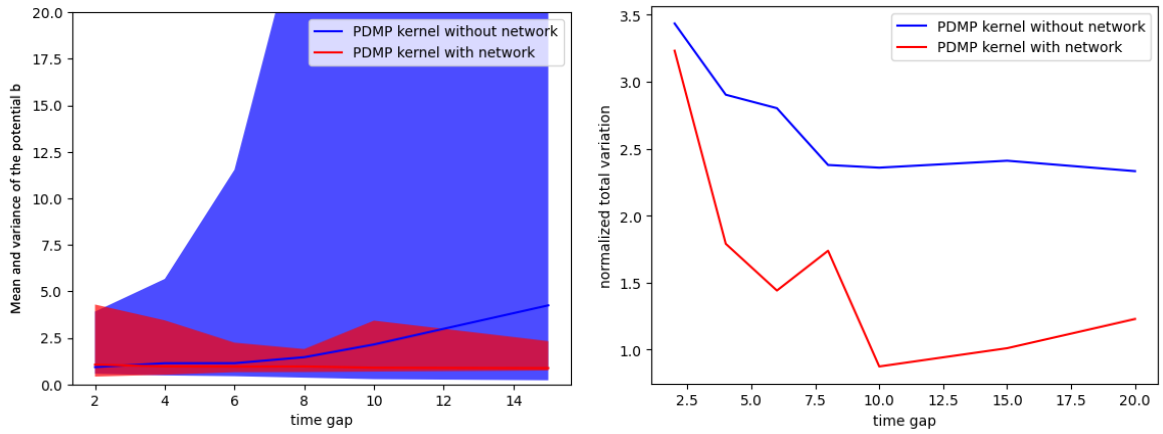
We also represent in Figure 6.3a the mean of the potential b_T obtained by the Sinkhorn algorithm, on the ending space G_N^n, as well as the $90th$ percentile of its variations. A mean value close to 1 associated to a small variation means by Proposition 28 that the optimal network is close to the one of the reference process (as it is the case when it is calibrated by the right network, in red), while a high variance means that the reference process is not calibrated by the right network (as it the case when it is calibrated by a null network, in blue).

We finally represent in Figure 6.4 the optimal burst rate function $k_{on,1}$ found by our method in the two cases (right and wrong calibration of the reference process), compared to the function used to simulate the data. We observe that the global form of the burst rate function is well reproduced in both cases, except in some areas of the gene expression space for the case of the wrong initial calibration. This may be due to the lack of data, preventing us from being able to identify the form of the function on these areas. This leads to incorrect evaluations when performing the inference part (using the relation (6.30)), and to a wrong network. Remark that this could be probably improved by identifying areas of the gene expression space for which we have no reliable information, which should be avoided from the grid G_N^n when solving this regression problem.

6.5 Discussion and limits

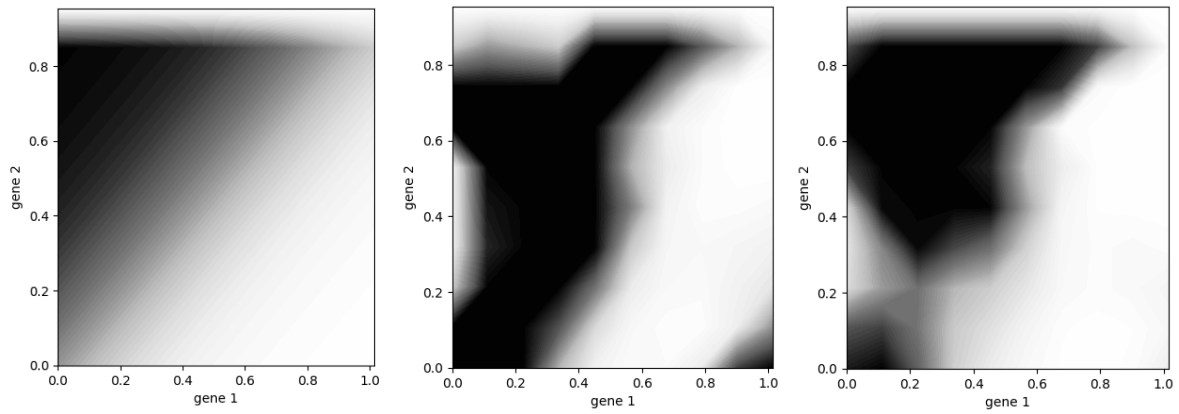
As emphasized in the introduction, the work that is presented here is still under construction. The results are promising, first mathematically, since we have obtained explicit formulas that allowed us to relate the different objects we manipulate, and to a lesser extent numerically. We have seen that we could construct a suitable entropic cost measuring the gap between the data and our model, and that it seemed possible to use this distance to infer the optimal parameters of the model with respect to the data. This work suffers nevertheless from several limitations.

First, we remark that the approach we developed in this Chapter, which transforms the inference problem in a set of regression problems of the form (6.30), specific to each gene, has similarities



(a) (b)

Figure 6.3: 6.3a Evolution of the entropic potential b computed with the Sinkhorn algorithm, as a function of the time gap between the two timepoints, when the reference kernel is computed with the right network (in red) and a null network (in blue). For the right network, we expect $b = \mathbb{1}$. 6.3b Distance of the network inferred with the formula (6.30) by using the kernels like the ones presented in Figures 6.4b (when the reference coupling is computed with a null network) in blue, and 6.4c (when the reference coupling is computed with the right network), estimated for the different values of the time gap between the measures.



(a) (b) (c)

Figure 6.4: Comparison of the burst rate functions of gene 1, $k_{on,1}$, 6.4a used for simulating the data, 6.4b estimated with the formula (6.28) when the reference coupling is computed with a null network and 6.4c estimated with the formula (6.28) when the reference coupling is computed with the right network. The levels of the colormap is the same for the three figures. For 6.4b and 6.4c, the reference coupling in the Schrödinger problem is computed for $T = 3h$.

with the one which had been developed in Chapter 3 (Section 4.4) for CARDAMOM. However, an important limit of this method is that that the sum has to be performed on the grid G_N^n leading to numerical intractability when the number of genes is high. This is difficult to overcome because we need, in order to be able to estimate the value of $k^{\theta}_{on,i}(x)$, to find for each cell $X \in S_0$ and for all gene i, at least one cell Y in the ending space such that $\forall j \neq i, \exists h, X_j = Y_j$, and $Y_i = X_i + h$. To our point of view, this method could be largely simplified by reducing the dimension of the problem, by using similar arguments as the ones developed for the algorithm CARDAMOM in Chapter 3. The gene expression space could be discretized in a grid much coarser than the one we used here, taking into account only the values leading to distinct modes for the promoters frequency, and the Schrödinger problem could be applied on the discrete measure describing the set of cells S_T on this simplified discrete space. Moreover, this could allow to overcome the fact that the current method requires protein measurements, which is critical as we recall that it is very difficult at the moment to observe proteins at the single-cell level experimentally [44]. Indeed, we could build from scRNA-seq data a coarse approximation of the proteins distributions on a discrete space, in a similar way that what we developed in Chapter 3, and then perform the Schrödinger problem analysis by considering these reduced observations. The mathematical link between these approximations and the results for the relative entropy developed in this section would nevertheless require further analyzes.

Second, the framework that we developed is restricted to the case of two timepoints, which prevents us from being able to compare the accuracy of the method to CARDAMOM for real datasets like the one used in Chapter 4. Developing this framework in the multi-marginal case, as it has been done in [47] when the reference process is an SDE with gradient drift, should be the subject of future works.

Finally, an important question that remains open for the moment is the sensibility of the method with respect to the choice of the jump kernel $\bar{q}$ used in the reference process $R^{\bar{q}}$. We do not expect a property similar to Theorem 9, which ensures in the diffusion case that the law of the process is correctly reconstructed by minimizing the entropy relatively to the Brownian motion. However, we claim that this is not as crucial as in [84, 47], because our method does not aim to reconstruct the characteristics of the process without prior knowledge. On the contrary, we have motivated our analysis by the need to test hypotheses and to study their limits regarding an experimental dataset. This work should be continued during my post-doctorate, which will be supervized by the last author of [84], Geoffrey Schiebinger.

Discussion and perspectives

Discussion

As introduced in Chapter 1 and at different stages of our manuscript, the notion of landscape is a powerful concept for representing the forces driving cellular differentiation and the resulting behavior of individual cells. The mathematical theory of dynamical systems, and of their perturbations, allow to define certain quantities that characterize the landscape associated to a stochastic system. In particular, the theory of Large deviations and our work of Part I allows to make the link between these two points of view for the mechanistic model 1.14. In this section, we highlight that the extended notions of *static* and *transitory* landscapes are particularly interesting for illustrating the framework of the different approaches that we developed in Parts II and III of this manuscript. In particular, we are going to detail how the landscape associated to the model 1.15 can appear as a generalization of the landscape associated to the phenomenological model presented at the end of Chapter 2, but also how the latter can question the modeling choices that have been made.
On one side, in this manuscript, we have made the hypothesis that the landscape was shaped by an underlying network. On the other side, some authors consider that perturbation of the system that induce differentiation may be seen as modification of the network. To our point of view, the word *Networks* itself generates ambiguity. It is essential to distinguish the *structure* and the *state* of the network. If one says that our network incorporates all critical nodes, then structure should not change. That of course would be very different if there are some " hidden variables " (i.e. genes the level of which are not measured), and which results in modifying the network structure. Say for example that at time 1, we observe A ← B and and time 2, we observe A B (no more interaction) this can only result from the effect of a hidden variable C that appears at time 2 and " cuts " the connection (i.e. a gene the product of which makes the promoter of B not accessible to A anymore). So if C is integrated, then the network structure does not change anymore (see Figure 6.5). In this point of view, assuming that the gene network does not vary with time consists in assuming that all the variables allowing to explain the data are observed.

However, even with these explanation, an ambiguity often remains about the parameters of the GRN: do they capture only chemical reactions involving only proteins and promoters, or do they take into account hidden variables, such as epigenetic marks ? In the first case, considering that we take into account in the model all the genes of interest (which is nevertheless a strong assumption), the network should therefore be constant regardless of everything else. However, as explained in the introduction and observed in Chapter 4 (see Figure 5), the GRN calibrating the burst rate functions is able to detect not only physical interactions between regulators and regulated genes, but also various epigenetic information or indirect effects. Following the same reasoning as in Figure 6.5, the processes characterizing these indirect effects, that are not observed nor directly modeled, could then affect the behavior of the model through a modification of the network. In that point of view, the hypothesis that the network does not change over time is then reasonable when we have to deal with "standard" single-cell data, but leads to errors that should be quantified.

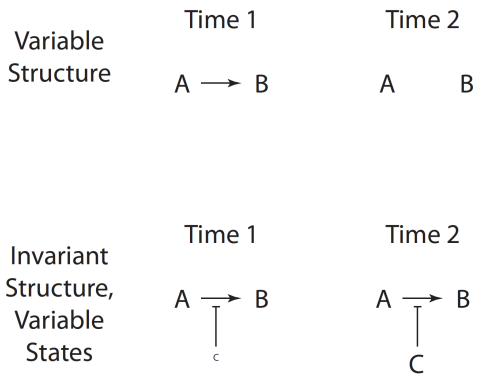


Figure 6.5: Action of a hidden variable C which modifies the structure of a network between A and B (figure courtesy of O. Gandrillon, personal communication).

Interestingly, that is precisely what allows the entropy minimization approach developed in Part III. The fact that this method provides time-dependent optimal jump kernels, even when the kernel of the reference process is constant, can be interpreted as a way of characterizing the influence of the hidden variables on the system, and the resulting time-dependent GRN as the modified states of the GRN under the action of these variables.
If it was possible to incorporate in the model every variables acting on the differentiation process, the model parameters would not depend on time. In that case, the landscape characterizing cellular differentiation is *statical*. The price to pay is nevertheless not negligible: this static landscape is potentially extremely complex, depending on all the parameters of this full model. This is the reason why, when modeling a biological process, we always make various assumptions and simplifications, in order to get a model, the complexity of which is proportional to the wealth of data available and ideally make the model parameters identifiable. In such situation, the model is affected by hidden variables and the landscape which characterizes differentiation is *dynamical*.

Without modeling assumptions, the best way to integrate this dynamical point of view would be to consider time-dependent parameters, as obtained with the method developed in Chapter 6. In order to link these modifications to interpretable phenomena, it would be necessary to consider "higher" models, taking into account various possible features such as knock-outs, methylation, cell-to-cell interactions, and to express the simplified model as a Markov *transitory* process from the global one, depending on these additional features. Considering the PDMP model (1.14), the natural way of doing that would be to make the burst rate functions $k_{on,i}$ (and eventually $k_{off,i}$) to depend on the hidden variables that we want to consider, denoted Y. This could be a way of interpreting in a general framework the dependence in time of these functions in in Chapter 6. Combining with a specific model for the dynamics of Y (which could be a diffusion, a discrete Markov chain, another PDMP process...) we could then understand the dynamical landscape as the landscape resulting from the coupling of the process on Y to the PDMP process (1.14) through these burst rate functions.

This provides for example a natural definition for knock-out potentials, fixing $Y = c$, c being a constant. Such potential corresponds to the transform $-\log \mu_{frozen}$, where μ_{frozen} is the stationary measure of the PDMP process (1.14) with burst rate functions $k_{on}(\cdot, c)$. We emphasize that it is exactly in that way that we have taken into account the effect of a stimulus in Part II, considering that Y denotes the state of an hidden gene acting on the network, which is set to 1 at time $t = 0$. This framework can also be used for characterizing the effect of a very slow

variable on the system. In the latter case, slow changes could be taken into account by modifying the value of c, at rare events characterizing jumps of the slow variable from one state to another. Interestingly, this is exactly the framework of the phenomenological process described in Chapter 2, and used in Chapter 3 for building the Gamma-mixture approximation used in the algorithm CARDAMOM. In this simplified model, the functions $k_{on,i}$ are considered to be constant within each basin, and then depends only of the slow variable characterizing the basins, which follows a Markov chain on the discrete space of metastable basins. As shown in Theorem 10, this model has the great advantage to have a known stationary distribution, which is a mixture of Beta distributions (and of Gamma distributions in the bursty regime).

This framework can also be extended to mean-field rate functions, by integrating a measure ν on the hidden variables Y. We take then:

$$k_{on,\nu}(x) = \int k_{on}(x,y)\nu(\mathrm{d}y). \tag{6.31}$$

This new rate provides a natural way to define a "mean-field" landscape $V_\nu(x) = -\log\big(\mu_\nu(x)\big)$, where μ_ν is the stationary measure of the PDMP process (1.14) with this new burst rate function. This framework should be particularly adapted for characterizing the effect of a fast variable on the system, which reaches its equilibrium knowing the observed variables X at each time.

With this method, we recover the modified rate functions $k^\alpha_{on,i}$ in Appendix A of Chapter 3, as the mean-field rate function associated to the phenomenological model described in Chapter 3, Section 3. Indeed, under the current notations, we recall that this function is defined for all i by:

$$k^\alpha_{on,i}(x) := \mathbb{E}(k_{y,i}|X=x) = \frac{\sum\limits_{y\in Y}\mu(y)k_{y,i}\prod_{j=1}^n \mathrm{Gamma}(k_{y_j},c_j)(x)}{\sum\limits_{y\in Y}\mu(y)\prod_{j=1}^n \mathrm{Gamma}(k_{y_j},c_j)(x)},$$

where μ denotes the marginal on the variable Y of the stationary distribution $\hat{u}$ of the phenomenological model. It corresponds exactly to the formula (6.31) when the density ν corresponds to the law of density $\hat{u}(y|X=x)$, with:

$$\hat{u}(dx,dy) = \mu(y)\prod_{j=1}^n \mathrm{Gamma}(k_{y_j},c_j)(x)dxdy.$$

We had also proved in Chapter 3 that in this case, the stationary distribution μ_ν is a Gamma-mixture, which corresponds to the stationary distribution of the phenomenological model. Note that this exact correspondence between the marginal (on the continuous variable X) of stationary distribution of a full model on (E,X,Y) and the one of the model on (E,X) with mean-field rate function defined by (6.31), when $\nu = \hat{u}(y|x)$, is not easily generalizable.

Finally, we summarize below how the few examples we have seen in this section can be formulated under the same formalism, and could be used to consider complex models and their relationship to simpler models. This list should be extended according to the needs, depending on the context, the underlying assumptions, reasonable approximations or information at disposal.

1. Taking $\nu(dy) = \delta_{y=c}$ leads to a notion of **knock-out landscape**, that allow to take into account disturbance that affects the whole trajectory, as knock-outs, or stimuli.

2. Taking $\nu(\mathrm{d}y) = \mu(\mathrm{d}\mid x)$ the conditional stationary distribution of the hidden variables Y_t knowing the observed variables X_t leads to a definition of **static landscape**, see the following section for an intuitive explanation about this terminology.

3. Taking $\nu(dy) = \mu_t(\mathrm{d} \mid x)$, with μ_t the joint temporal distribution of (X_t, Y_t), corresponds to a natural definition for a **transitory landscape**.

The last case is the most interesting, as it could allow to directly link the time-dependent burst rate functions obtained from entropy minimization problems to the characteristics of the hidden processes. The two previous cases are nevertheless to be kept in mind when considering GRN models only, as they highlight the approximation that may be not evidenced when considering a model with specific burst rate functions. This is for example the case for the mechanistic model of the k_{on} functions developed in [38], that we used in Chapter 2. We believe that this general point of view could pave the way for a methodology able to extend the model in a tractable way, *i.e* adapted both to our knowledge and the nature of the available data.

Conclusion

The goal of the thesis was to develop a general framework for describing the landscape of cellular differentiation under the hypothesis that it is shaped by an underlying GRN, and to reconstruct it from single-cell data, when the latter is set up by biologically interpretable parameters. We started from a mechanistic model developed in a previous thesis [36], which had been shown to accurately reproduce the experimentally-observed distribution of gene expression products. We adopted a different approach, investigating mathematical tools related to statistical physics like Large deviations theory and optimal transport, that we developed for such non-diffusive process. In terms of landscape, we used results from [25, 14] to build an approximate landscape by computing exactly the transitions rates between the different cell types emerging from a GRN, seen as *macrostates*. This led us to propose a phenomenological model, the distribution of which is close to a mixture of Gamma distributions specific to each cell type. This simplified model, combined to the analytical description of the approximate landscape as a function of the GRN, has allowed us to build an efficient algorithm for reconstructing a most-likely GRN from time-course series of gene expression datasets. Although this method contains some heuristic arguments, and that the identifiability is not ensured, we have shown that it is able to reproduce accurately the characteristics of the approximate landscape inferred from experimental observations, and then the main characteristics of gene expression dynamics. Thus, contrary to methods for reconstructing a landscape that are purely statistical [73, 81], or based on diffusion approximation [84, 13], our algorithm starts from a complex mechanistic model directly able to reproduce the characteristics of single-cell data without any addition of artificial noise, and leads to a simple statistical model where all the parameters are directly interpretable from the original model. This is also very different from most of the methods that are used in GRN inference, which do not consider variability in a mechanistic way, making the inferred interactions difficult to relate to an underlying landscape of cell differentiation. One of the great advantage of our work is that it explicitly combines GRN inference and landscape reconstruction in a single task.

Once we obtained a calibrated mechanistic model from single-cell datasets, the challenge that naturally arose was to evaluate both the accuracy of the calibration and the model. Indeed, even well calibrated, it may appear that the model is not adapted to the data used for the calibration, and we should have an appropriate distance for estimating whether it is the case or not. Following our statistical physics point of view on cellular differentiation, we used the relative entropy for quantifying the difference between the data and the calibrated model: due to Sanov's theorem, the Schrödinger problem indeed appears as a natural tool for this task. Using theoretical results of [52], we thus proposed a method for solving the Schrödinger problem when the reference process is the model calibrated by CARDAMOM, and used the solution for estimating the accuracy of both the calibration and the model. We observed first that this approach seems to be well suited for estimating the accuracy of a model with respect to the data, and interpreting its successes and limitations, but less for reconstructing the underlying

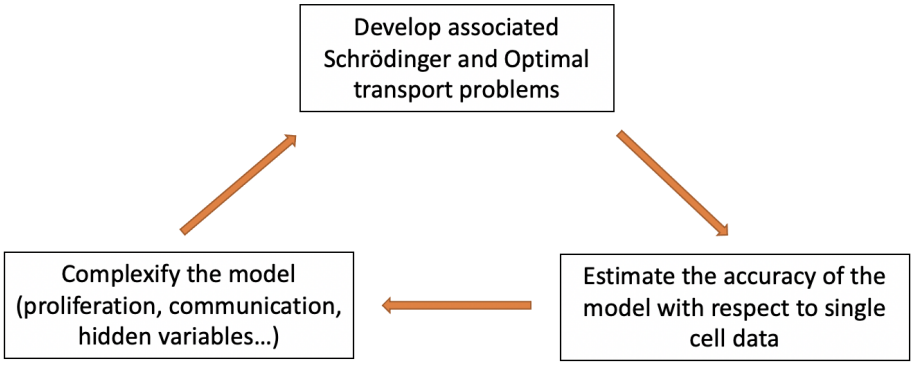


Figure 6.6: Conceptual framework for using entropy minimization problems (or optimal transport theory at the limit) integrating new features to the models and testing hypotheses against single-cell datasets.

landscape when we have no accurate prior knowledge on the parameters that shape it. Note that does not contradict the works developed in [84, 47, 107]: our approach can rather be seen as the preliminaries to a mechanistic version of them. Indeed, the authors of these works deal with high-dimensional data, for which a mechanistic approach would be very difficult: it is then natural for them to use a diffusion approximation of the process driving gene expression dynamics, for which properties specific to the Brownian motion ensure that efficient computational methods based on optimal transport can be used to reconstruct a landscape. To our point of view, this non-mechanistic approach should be considered not as a complete tool for understanding the whole landscape complexity, but as a pre-processing step of the data allowing to identify some characteristics of the landscape (for example, subsets of genes or subspaces of interest in the gene expression space...), in order to apply in a second step mechanistic-based methods.

As illustrated in Figure 6.6, we believe that the theory of Schrödinger problems (and optimal transport at the weak noise limit) could be efficiently used in a general framework as a validation step of a mechanistic model when compared to single-cells datasets. The dependence in time of the optimal characteristics of the model associated to the solution of the Schrödinger problem could also serves as an help for modeling choices, like the way of taking into account additional hidden processes in the burst rate functions, as discussed previously. A main advantage of this framework is its adaptability. Indeed, it has the great advantage to allow the integration of complex and multi-faceted information in the data, as it has been initiated for proliferation or lineage tracing estimation in [47] and [28] respectively, when the reference process is a stochastic diffusion. In the era of multiomics data [49, 90], this makes this theory very promising. In particular, very recent developments in the Schrödinger problem theory concerning stochastic processes with proliferation [7] or mean field Schrödinger problem [6] could pave the way for developing such mathematical framework allowing to consider proliferation or cell-to-cell communication in a mechanistic way. However, analytical and statistical methods adapted for these new extended models, as the ones developed in Parts I and II of this manuscript for the GRN model (1.14), should also be developed in parallel for calibration purposes.

Take-home message

The message of this thesis is double. First, we highlighted the benefits of adopting a probabilistic point of view on gene expression dynamics, demonstrating the possibility of developing an approach similar to what is done in statistical mechanics with models of cell differentiation taking into account the complexity of molecular mechanisms and features like transcriptional bursting. This is an important step in the direction of what advocated the authors of [91], that

"more bottom-up modelling of single-cell data were needed to help attain a deeper understanding of single-cell systems biology", or more recently the authors of [34] that "we are unaware of any trajectory inference methods that explicitly parameterize the underlying stochastic model using the Chemical Master Equation". In particular, we developed in Part II a method which uses metastability and Large deviations analysis for transforming the reverse-engineering problem of a probabilistic model on a set of regression problems closely related to the learning of a neural network. This approach has the great advantage of combining a mechanistic point of view on gene expression with the numerical power of machine learning algorithms (but it is worth noticing that it would probably loose its interpretability in case of too highly dimensional problems). As this method has been developed under the hypothesis that simple rate functions characterize the dynamics of the model, the neural network associated to the regression problems are simple one-layer perceptron, but there would be no obstacle to extend this same method to more complex functions (and thus networks) if the data seemed to require it.
Second, our work shed a new light on the complex notions of static and transitory landscape. We have seen with the phenomenological model developed in Chapter 2 and used afterwards that splitting a static landscape into simpler transitory landscapes can allow the original landscape to be reconstructed piece by piece. This corresponds to a mathematical description in terms of landscape of the notion of transitory states defined in Moris et al. [66] for characterizing cell decision mechanisms during differentiation. Moreover the method developed in Chapter 6, which allows to build an optimal time-dependent kernel from a reference process and partial observations of a system, provides a rational for understanding the limits of a static landscape point of view for understanding the system. It could be used to initiate a modeling work in order to integrate this static landscape associated to the reference process into a more complex landscape that would better explains the data. Doing this, we use the Schrödinger problem not as a machine learning tool for inferring trajectories from high-dimensional data, as in [84], but as a way of refining and complexifying the landscape of cell differentiation. To this end, it would be necessary to develop modeling approaches like the one outlined in the Discussion, to link the dynamical landscape associated to the solution of the Schrödinger problem with a bigger static landscape integrating new dynamical processes interacting with the GRN.

Bibliography

[1] A. Aalto, L. Viitasaari, P. Ilmonen, L. Mombaerts, and J. Gonçalves. "Gene regulatory network inference from sparsely sampled noisy data". In: *Nature communications* 11.1 (2020), pp. 1–9.

[2] S. Adams. "Large deviations for stochastic processes". In: *Report Eurandom* 2012025 (2012).

[3] K. Akers and T.M. Murali. "Gene regulatory network inference in single cell biology". In: *Current Opinion in Systems Biology* (2021).

[4] C. Albayrak, C. A. Jordi, C. Zechner, J. Lin, C. A. Bichsel, M. Khammash, and S. Tay. "Digital Quantification of Proteins and mRNA in Single Mammalian Cells". In: *Molecular Cell* 61 (2016), pp. 914–924.

[5] Pi-C. Aubin-Frankowski and J-P. Vert. "Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference". In: *Bioinformatics* 36.18 (2020), pp. 4774–4780.

[6] J. Backhoff, G. Conforti, I. Gentil, and C. Léonard. "The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities". In: *Probability Theory and Related Fields* 178.1 (2020), pp. 475–530.

[7] A. Baradat and H. Lavenant. "Regularized unbalanced optimal transport as entropy minimization with respect to branching Brownian motion". In: *arXiv preprint arXiv:2111.01666* (2021).

[8] A. Baradat and E. Ventre. "Convergence of the Sinkhorn algorithm when the Schrödinger problem has no solution". In: *arXiv preprint arXiv:2111.01666* (2022).

[9] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. "Qualitative properties of certain piecewise deterministic Markov processes". In: *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 51 (2015), pp. 1040–1075.

[10] J-D. Benamou and Y. Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numerische Mathematik* 84.3 (2000), pp. 375–393.

[11] M. Bizzarri, M. G. Masiello, A. Giuliani, and A. Cucina. "Gravity Constraints Drive Biological Systems Toward Specific Organization Patterns: Commitment of cell specification is constrained by physical cues". In: *Bioessays* (2017).

[12] A. Bonnaffoux, U. Herbach, A. Richard, A. Guillemin, S. Gonin-Giraud, P.-A. Gros, and O. Gandrillon. "WASABI: a dynamic iterative framework for gene regulatory network inference". In: *BMC Bioinformatics* 20 (2019), p. 220.

[13] R. D. Brackston, A. Wynn, and M. P. Stumpf. "Construction of quasipotentials for stochastic dynamical systems: An optimization approach". In: *Physical Review E* 98.2 (2018), p. 022136.

[14] P. Bressloff and O. Faugeras. "On the Hamiltonian structure of large deviations in stochastic hybrid systems". In: *Journal of Statistical Mechanics: Theory and Experiment* 2017.3 (2017), p. 033206.

[15] T. E Chan, M. P H Stumpf, and A. C Babtie. "Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures". In: *Cell Systems* 5 (2017).

[16] S. Chen and J. C. Mar. "Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data". In: *BMC Bioinformatics* 19 (2018).

[17] L. Chizat, G. Peyré, B. Schmitzer, and F-X. Vialard. "Scaling algorithms for unbalanced optimal transport problems". In: *Mathematics of Computation* 87.314 (2018), pp. 2563–2609.

[18] H. Clevers, S. Rafelski, M. B. Elowitz, A. Klein, A. Shendure, C. Trapnell, E. Lein, E. Lundberg, M. Uhlen, A. Martinez Arias, J. R. Sanes, P. C. Blainey, J. Eberwine, J. Kim, and J. C. Loven. "What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism?" In: *Cell Systems* 4 (2017), pp. 255–259.

[19] M. Coomer, L. Ham, and M. Stumpf. "Shaping the epigenetic landscape: Complexities and consequences". In: *bioRxiv* (2020).

[20] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. "Convergence of stochastic gene networks to hybrid piecewise deterministic processes". In: *The Annals of Applied Probability* 22 (2012), pp. 1822–1859.

[21] J. Dattani and M. Barahona. "Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization". In: *Journal of the Royal Society, Interface* 14 (2017).

[22] M. Davis. "Jump Processes and Their Martingales". In: *Preprint series: Pure mathematics http://urn. nb. no/URN: NBN: no-8076* (1991).

[23] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. 1998.

[24] P. Dibaeinia and S. Sinha. "SERGIO: a single-cell expression simulator guided by gene regulatory networks". In: *Cell systems* (2020).

[25] A. Faggionato, D. Gabrielli, and M. Ribezzi Crivellari. "Non-equilibrium Thermodynamics of Piecewise Deterministic Markov Processes". In: *Journal of Statistical Physics* 137 (2009), pp. 259–304.

[26] J. Feng and T. Kurtz. *Large deviations for stochastic processes*. 131. American Mathematical Soc., 2006.

[27] H. Föllmer. "Random fields and diffusion processes". In: *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. Springer, 1988, pp. 101–203.

[28] A. Forrow and G. Schiebinger. "LineageOT is a unified framework for lineage tracing and trajectory inference". In: *Nature communications* 12 (2021).

[29] M.I. Freidlin and A. Wentzell. "Random perturbations". In: *Random perturbations of dynamical systems*. Springer, 1998, pp. 15–43.

[30] N. Friedman, L. Cai, and X S. Xie. "Linking stochastic dynamics to population distribution: an analytical framework of gene expression". In: *Phys Rev Lett* 97 (2006).

[31] N.P. Gao, O. Gandrillon, A. Páldi, U. Herbach, and R. Gunawan. "Universality of cell differentiation trajectories revealed by a reconstruction of transcriptional uncertainty landscapes from single-cell transcriptomic data". In: *bioRxiv* (2020).

[32] I. Gentil, C. Léonard, and L. Ripani. "About the analogy between optimal transport and minimal entropy". In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 26. 3. 2017, pp. 569–600.

[33] E. Giusti and GH. Williams. *Minimal surfaces and functions of bounded variation*. Vol. 80. Springer, 1984.

[34] G. Gorin, M. Fang, T. Chari, and L. Pachter. "RNA velocity unraveled". In: *bioRxiv* (2022).

[35] U. Herbach. "Gene regulatory network inference from single-cell data using a self-consistent proteomic field". In: *arXiv* 2109.14888 (2021).

[36] U. Herbach. "Modélisation stochastique de l'expression des gènes et inférence de réseaux de régulation". PhD thesis. Université de Lyon, 2018.

[37] U. Herbach. "Stochastic gene expression with a multistate promoter: breaking down exact distributions". In: *SIAM Journal on Applied Mathematics* 79 (2019), pp. 1007–1029.

[38] U. Herbach, A. Bonnaffoux, T. Espinasse, and O. Gandrillon. "Inferring gene regulatory networks from single-cell data: a mechanistic approach". In: *BMC Systems Biology* 11 (2017), p. 105.

[39] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber. "Cell fates as high-dimensional attractor states of a complex gene regulatory network". In: *Physical review letters* 94.12 (2005), p. 128701.

[40] V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts. "Inferring regulatory networks from expression data using tree-based methods". In: *PLOS One* 5 (2010).

[41] V. A. Huynh-Thu and G. Sanguinetti. "Gene Regulatory Network Inference: An Introductory Survey". In: *Methods Mol Biol* 1883 (2019), pp. 1–23.

[42] B-R. Hyun, J. McElwee, and P. Soloway. "Single molecule and single cell epigenomics". In: *Methods* 72 (2015), pp. 41–50.

[43] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. "Stochasticity in gene expression: from theories to phenotypes". In: *Nat Rev Genet* 6 (2005), pp. 451–464.

[44] R. Kelly. "Single-cell proteomics: progress and prospects". In: *Molecular & Cellular Proteomics* 19.11 (2020), pp. 1739–1748.

[45] M. S. Ko, H. Nakauchi, and N. Takahashi. "The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates." In: *The EMBO journal* 9 (1990), pp. 2835–2842.

[46] B. Laforge, D. Guez, M. Martinez, and J. Kupiec. "Modeling embryogenesis and cancer: an approach based on an equilibrium between the autostabilization of stochastic gene expression and the interdependence of cells for proliferation". In: *Progress in biophysics and molecular biology* 89.1 (2005), pp. 93–120.

[47] H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger. "Towards a mathematical theory of trajectory inference". In: *arXiv preprint arXiv:2102.09204* (2021).

[48] C. Lawrence. "Partial Differential Equations". In: *Graduate Studies in Mathematics* 19 (2010), pp. 333–339.

[49] J. Lee, DY. Hyeon, and D. Hwang. "Single-cell multiomics: technologies and data analysis methods". In: *Experimental & Molecular Medicine* 52.9 (2020), pp. 1428–1442.

[50] C. Léonard. "From the Schrödinger problem to the Monge–Kantorovich problem". In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920.

[51] C. Léonard. "Girsanov theory under a finite entropy condition". In: *Séminaire de Probabilités XLIV*. Springer, 2012, pp. 429–465.

[52] C. Léonard. “Some properties of path measures”. In: *Séminaire de Probabilités XLVI*. Springer, 2014, pp. 207–230.

[53] G-W. Li and X. S. Xie. “Central dogma at the single-molecule level in living cells”. In: *Nature* 475.7356 (2011), pp. 308–315.

[54] Y. T. Lin and T. Galla. “Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models”. In: *J. R. Soc. Interface* 13 (2016).

[55] C. Lv, X. Li, F. Li, and T. Li. “Constructing the energy landscape for genetic switching system driven by intrinsic noise”. In: *PLoS one* 9.2 (2014).

[56] M. C Mackey, M. Tyran-Kamińska, and R. Yvinec. “Molecular distributions in gene regulatory dynamics”. In: *Journal of Theoretical Biology* 274 (2011).

[57] F. Malrieu. “Some simple but challenging Markov processes”. In: *Annales de la Faculté de Sciences de Toulouse* 24 (2015), pp. 857–883.

[58] J. C. Mar. “The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond”. In: *Biophysical Reviews* 11 (2019), pp. 89–94.

[59] M. Mariani. “A Gamma-convergence approach to large deviations”. In: *arXiv preprint arXiv:1204.0640* (2012).

[60] H. Matsumoto, H. Kiryu, C. Furusawa, M. S. Ko, S. B. Ko, N. Gouda, T. Hayashi, and I. Nikaido. “SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation”. In: *Bioinformatics* (2017).

[61] R. McCann. “A convexity principle for interacting gases”. In: *Advances in mathematics* 128.1 (1997), pp. 153–179.

[62] L. McInnes, J. Healy, and J. Melville. “UMAP: uniform manifold approximation and projection for dimension reduction”. In: *arXiv* 1802.03426 (2020).

[63] A. McKenna, G. Findlay, J. Gagnon, M. Horwitz, A. Schier, and J. Shendure. “Whole-organism lineage tracing by combinatorial and cumulative genome editing”. In: *Science* 353.6298 (2016), aaf7907.

[64] J. Merkin, C. Russell, P. Chen, and C. Burge. “Evolutionary dynamics of gene and isoform regulation in Mammalian tissues”. In: *Science* 338.6114 (2012), pp. 1593–1599.

[65] A. Mizeranschi, H. Zheng, P. Thompson, and W. Dubitzky. “Evaluating a common semi-mechanistic mathematical model of gene-regulatory networks”. In: *BMC Systems Biology* 9 (2015), pp. 1–12.

[66] N. Moris, C. Pina, and A. M. Arias. “Transition states and cell fate decisions in epigenetic landscapes”. In: *Nature Reviews Genetics* 17 (2016).

[67] S. A. Morris. “The evolving concept of cell identity in the single cell era”. In: *Development* 146.12 (2019), pp. 1–5.

[68] H. Nguyen, D. Tran, B. Tran, B. Pehlivan, and T. Nguyen. “A comprehensive survey of regulatory network inference methods using single-cell RNA sequencing data”. In: *Briefings in Bioinformatics* (2020).

[69] D. Nicolas, NE. Phillips, and F. Naef. “What shapes eukaryotic transcriptional bursting?” In: *Mol Biosyst* 13.7 (2017), pp. 1280–1290.

[70] M. Nutz. “Introduction to Entropic Optimal Transport”. In: (2021).

[71] H. Ochiai, T. Sugawara, T. Sakuma, and T. Yamamoto. “Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells”. In: *Scientific reports* 4 (2014), pp. 1–9.

[72] N. Papili Gao, S. M. M. Ud-Dean, O. Gandrillon, and R. Gunawan. "SINCERITIES: Inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles". In: *Bioinformatics* (2017).

[73] P. Pearce, F. Woodhouse, A. Forrow, A. Kelly, H. Kusumaatmaja, and J. Dunkel. "Learning dynamical information from static protein and sequencing data". In: *Nature communications* 10.1 (2019), pp. 1–8.

[74] J. Peccoud and B. Ycart. "Markovian modeling of gene-product synthesis". In: *Theoretical population biology* 48 (1995), pp. 222–234.

[75] G. Peyré and M. Cuturi. "Computational optimal transport: With applications to data science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[76] X. Qiu, A. Rahimzamani, L. Wang, B. Ren, Q. Mao, T. Durham, J.L. McFaline-Figueroa, L. Saunders, C. Trapnell, and S. Kannan. "Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe". In: *Cell Systems* 10 (2020), pp. 1–10.

[77] A. Richard, L. Boullu, U. Herbach, A. Bonnafoux, V. Morin, E. Vallin, A. Guillemin, N. Papili Gao, R. Gunawan, J. Cosette, O. Arnaud, J. J. Kupiec, T. Espinasse, S. Gonin-Giraud, and O. Gandrillon. "Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process". In: *PLoS Biol* 14 (2016), e1002585.

[78] L. Ripani. "Le problème de Schrödinger et ses liens avec le transport optimal et les inégalités fonctionnelles". PhD thesis. Université de Lyon, 2017.

[79] J. Rodriguez and DR. Larson. "Transcription in Living Cells: Molecular Mechanisms of Bursting". In: *Annu Rev Biochem* 89 (2020), pp. 189–212.

[80] R. Rudnicki and A. Tomski. "On a stochastic gene expression with pre-mRNA, mRNA and protein contribution". In: *Journal of Theoretical Biology* 387 (2015), pp. 54–67.

[81] M. Sáez, R. Blassberg, E. Camacho-Aguilar, E. Siggia, D. Rand, and J. Briscoe. "Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions". In: *Cell Systems* 13.1 (2022), pp. 12–28.

[82] I.N. Sanov. *On the probability of large deviations of random variables.* Tech. rep. North Carolina State University. Dept. of Statistics, 1958.

[83] A. K Sarkar and M. Stephens. "Separating measurement and expression models clarifies confusion in single cell RNA-seq analysis". In: *BioRxiv* (2020).

[84] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, and E. S. Lander. "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming". In: *Cell* 176 (2019), 928–943 e22.

[85] E. Schrödinger. "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique". In: *Annales de l'institut Henri Poincaré.* Vol. 2. 4. 1932, pp. 269–310.

[86] B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. "Global quantification of mammalian gene expression control". In: *Nature* 473 (2011), pp. 337–42.

[87] V. Shahrezaei and P. S. Swain. "Analytical distributions for stochastic gene expression". In: *PNAS* 105 (2008), pp. 17256–17261.

[88] S. Sinha, A. Satpathy, W. Zhou, H. Ji, J. Stratton, A. Jaffer, N. Bahlis, S. Morrissy, and J. Biernaskie. "Profiling chromatin accessibility at single-cell resolution". In: *Genomics, Proteomics & Bioinformatics* 19.2 (2021), pp. 172–190.

[89] K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and S. Dudoit. "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics". In: *BMC Genomics* 19 (2018), p. 477.

[90] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika. "Multi-omics data integration, interpretation, and its application". In: *Bioinformatics and biology insights* 14 (2020), p. 1177932219899051.

[91] A. Teschendorff and A. Feinberg. "Statistical mechanics meets single-cell biology". In: *Nature Reviews Genetics* 22.7 (2021), pp. 459–476.

[92] H. Touchette. "The large deviation approach to statistical mechanics". In: *Physics Reports* 478.1-3 (2009), pp. 1–69.

[93] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, N. Morse M.and Lennon, K. Livak, T. Mikkelsen, and J. Rinn. "Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions". In: *Nature biotechnology* 32.4 (2014), p. 381.

[94] E. Tunnacliffe and J. Chubb. "What Is a Transcriptional Burst?" In: *Trends Genet* 36.4 (2020), pp. 288–297.

[95] S. Ulianov, K. Tachibana-Konwalski, and S. Razin. "Single-cell Hi-C bridges microscopy and genome-wide sequencing approaches to study 3D chromatin organization". In: *BioEssays* 39.10 (2017), p. 1700104.

[96] E. Ventre. "Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics". In: *In Silico Biology* 14 (2021), pp. 89–113.

[97] E. Ventre, T. Espinasse, C.-E. Bréhier, V. Calvez, T. Lepoutre, and O. Gandrillon. "Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastabilty". In: *Journal of Mathematical Biology* 83 (2021), pp. 1–63.

[98] Elias Ventre, Ulysse Herbach, Thibault Espinasse, Gérard Benoit, and Olivier Gandrillon. "One model fits all: combining inference and simulation of gene regulatory networks". In: *bioRxiv* (2022).

[99] L. Vistain and S. Tay. "Single-cell proteomics". In: *Trends in biochemical sciences* 46.8 (2021), pp. 661–672.

[100] C.H. Waddington. *The strategy of the genes*. Routledge, 1957.

[101] J. Wang, L. Xu, E. Wang, and S. Huang. "The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation". In: *Biophysical journal* 99.1 (2010), pp. 29–39.

[102] J. Wang, K. Zhang, L. Xu, and E. Wang. "Quantifying the Waddington landscape and biological paths for development and differentiation". In: *Proceedings of the National Academy of Sciences* 108.20 (2011), pp. 8257–8262.

[103] Y.X.R. Wang and H. Huang. "Review on statistical methods for gene network reconstruction using expression data". In: *Journal of Theoretical Biology* 362 (2014), pp. 53–61.

[104] C. Weinreb, S. Wolock, B. Tusi, M. Socolovsky, and A. Klein. "Fundamental limits on dynamic inference from single-cell snapshots". In: *Proceedings of the National Academy of Sciences* 115.10 (2018), E2467–E2476.

[105] R. Yvinec, C. Zhuge, J. Lei, and M. Mackey. "Adiabatic reduction of a model of stochastic gene expression with jump Markov process". In: *Journal of mathematical biology* 68.5 (2014), pp. 1051–1070.

[106] B. Zhang and P. G. Wolynes. “Stem cell differentiation as a many-body problem”. In: *PNAS* 111 (2014), pp. 10185–10190.

[107] S. Zhang, A. Afanassiev, L. Greenstreet, T. Matsumoto, and G. Schiebinger. “Optimal transport analysis reveals trajectories in steady-state systems”. In: *PLoS computational biology* 17.12 (2021), e1009466.

[108] J. X. Zhou, MDS. Aliyu, E. Aurell, and S. Huang. “Quasi-potential landscape in complex multi-stable systems”. In: *Journal of the Royal Society Interface* 9.77 (2012), pp. 3539–3553.

[109] P. Zhou and T. Li. “Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, A-type integral and beyond”. In: *The Journal of chemical physics* 144.9 (2016), p. 094109.

[110] S. Zreika, C. Fourneaux, E. Vallin, L. Modolo, R. Seraphin, A. Moussy, E. Ventre, M. Bouvier, A. Ozier-Lafontaine, A. Bonnaffoux, F. Picard, Gandrillon; O., and S. Giraud. “Evidence for close molecular proximity between reverting and undifferentiated cells”. In: *bioRxiv* (2022).