

On the Mathematics of RNA Velocity I: Theoretical Analysis

Tiejun Li^{1,*}, Jifan Shi^{2,*}, Yichong Wu^{1,*} and Peijie Zhou^{3,*}

¹ LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China.

² International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo 113-0033, Japan.

³ Department of Mathematics, University of California, Irvine, Irvine, CA 92697, USA.

Received 21 September 2020; Accepted 29 October 2020

Abstract. The RNA velocity provides a new avenue to study the stemness and lineage of cells in the development in scRNA-seq data analysis. Some promising extensions of it are proposed and the community is experiencing a fast developing period. However, in this stage, it is of prime importance to revisit the whole process of RNA velocity analysis from the mathematical point of view, which will help to understand the rationale and drawbacks of different proposals. The current paper is devoted to this purpose. We present a thorough mathematical study on the RNA velocity model from dynamics to downstream data analysis. We derived the analytical solution of the RNA velocity model from both deterministic and stochastic point of view. We presented the parameter inference framework based on the maximum likelihood estimate. We also derived the continuum limit of different downstream analysis methods, which provides insights on the construction of transition probability matrix, root and ending-cells identification, and the development routes finding. The overall analysis aims at providing a mathematical basis for more advanced design and development of RNA velocity type methods in the future.

AMS subject classifications: 60J28, 62P10, 92B15

Key words: RNA velocity, stochastic model, continuum limit, kNN density estimate.

1 Introduction

Single-cell RNA sequencing (scRNA-seq) is a rapid maturing technique, which makes the elaborate study of biological processes in the single cell resolution possible [50, 58].

*Corresponding author. *Email addresses:* tieli@pku.edu.cn (T. Li), shijifan@ircn.jp (J. Shi), wuyichongyt@pku.edu.cn (Y. Wu), peijiez1@uci.edu (P. Zhou)

The rich and diverse scRNA-seq datasets are revealing to us the mysteries of stem cell differentiation [54], heterogeneity in multicellular organisms [25], cancer cell dissection [8,37,64], drug discovery [22,59], etc. Every year, a swarm of analysis tools are produced by researchers all over the world [42, 62]. Some popular choices include the clustering tools [6, 28], trajectory inference tools [21, 40, 42, 47, 51, 53], and energy landscape tools [26,44,46,63], etc.

The characterization of stemness and lineage of the cells is a fundamental question in developmental biology. Although some practical indices, such as the signalling entropy and Markov chain entropy [47, 51], etc., are proposed to quantify the stemness of different cells in the scRNA-seq data analysis, they are more or less heuristic in nature. Recently, another promising method, the RNA velocity [29], was proposed to address this issue based upon the fact that the nascent (unspliced) and mature (spliced) mRNA can be distinguished in common single-cell RNA-seq protocols, such as SMART-seq2 [38], Drop-seq [32] and 10X genomics [66]. Thus, the relative abundance of unspliced and spliced mRNA are utilized to infer the velocity of each cell in the spliced mRNA abundance space, and predict the tendency of transition from one cell to another according to the RNA velocity model [29]. Improved methods in kinetic modeling, parameter inference and downstream analysis have been subsequently proposed [3,41], showing the potential of RNA velocity to quantify the stemness of cells in a rational way.

Despite the fruitful results and promising applications of RNA velocity, it is of prime importance to understand the rationale underlying the algorithm design, as well as the subtle differences between different proposals from mathematical point of view. For instance, when constructing the cell-cell stochastic transition probability matrix from RNA velocity, La Mano et al. [29] and Qiu et al. [41] used the correlation scheme in the velocity kernel, while the cosine scheme was proposed in [3]. In the recent version of dynamo package [39], a scheme with local kernels [4] of diffusion was also utilized. In spite of their intuitive plausibility, the theoretical implications of different kernels demands further investigation. In addition, a tracking strategy of root and ending cells has been applied based on forward and backward diffusions [3,29], whose theoretical basis remains to be established. Resolution of these puzzles based on a formal mathematical study will not only shed light on these theoretical problems, but also lead to a deeper comprehension of the RNA velocity and inspire further rational design of more delicate RNA velocity models. The current paper is devoted to this purpose.

In this work, we will present a thorough mathematical study on the whole process of RNA velocity model from kinetic model derivation, parameter inference algorithm to the downstream dynamical analysis. Our analysis will contribute insights toward several fundamental questions regarding RNA velocity and relevant downstream analysis, including:

- How to derive the deterministic and stochastic kinetic models of RNA velocity, and find analytical solutions?
- How to build the maximum likelihood estimator (MLE) of the parameters, built on

the exact solution of stochastic RNA velocity model?

- How can the discrete cellular transition dynamics inferred from RNA velocity be rigorously associated the continuous dynamical system model in cell-fate decision?
- What is the essential difference between the choice of correlation, cosine or inner-product scheme in the velocity kernel for the cellular transition matrix?
- What is the implication to replace the Gaussian scheme with k-nearest neighbor (kNN) scheme in the diffusion kernel?
- Why is the backward and forward diffusion strategy effective in detecting root and ending cells of development?
- How to rationally construct developmental trajectories based on RNA velocity with mathematical theory, beyond illustrating arrows and streamlines in the reduced dimension space?

We will focus on the formal mathematical analysis in the current paper and leave detailed computational comparisons and improvements in the continued publication [30]. To the best knowledge of the authors, this is the first attempt on studying the mathematics of RNA velocity in a complete manner. We hope it will provide a mathematical basis for further development of RNA velocity type methods in the future.

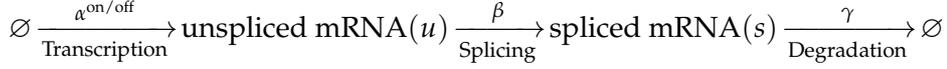
The rest of the paper is organized as follows. In Section 2, we show the mathematical derivations of both deterministic and stochastic kinetic models of RNA velocity, and derive the associated analytical solutions. In Section 3 we will revisit the existing algorithms to infer parameters in RNA velocity models, and present a novel maximum likelihood estimation of parameters originated from the exact solution of stochastic models. In Section 4 we focus on the dynamical system analysis based on RNA velocity, deriving the continuum limit of discrete transition probabilities with various kernels, demonstrating the mathematical rationale for the existing strategy of root/ending cells detection, and providing a new method to construct development trajectories with RNA velocity through the well-established transition path theory. Finally we give the conclusion and discussions in Section 5. Some analysis details, such as the almost sure type convergence order of the kNN radius, are left in the Appendix.

2 Models of RNA velocity

The key point of the RNA velocity model of single cells is that one can identify the abundance of the nascent (unspliced) and mature (spliced) mRNA from the single-cell RNA-seq data, which provides the information on the time-dependent evolution rate of the mRNA abundance by incorporating appropriate dynamical models.

Denote by u and s the abundance of the unspliced and spliced mRNA, respectively. In its simplest form, the transcriptional dynamics of the mRNA velocity can be described by

Table 1: Schematics of the dynamics for the RNA velocity model.



the reaction pathways shown in Table 1 [29]. We assume the production of u is dictated by a transcriptional induction or repression with parameter α^{on} or α^{off} , respectively. The unspliced mRNA, u , is then transformed into the spliced form with rate β , and the spliced mRNA is eventually degraded with rate γ . We remark that the above statement must be understood for single gene, i.e. the parameters $(\alpha^{\text{on/off}}, \beta, \gamma)$ should be replaced by $(\alpha_g^{\text{on/off}}, \beta_g, \gamma_g)$ when we consider the dynamics for a specific gene g . But we will omit the g -dependence of the parameters for brevity if not necessary. In the current stage, we assume that there are no interactions among different genes.

The task in this section is to study the explicit solution and related analytical properties of the forward mRNA velocity model in both deterministic and stochastic forms, given the dynamical parameters $(\alpha^{\text{on/off}}, \beta, \gamma)$.

2.1 Deterministic model

The deterministic model of the reaction dynamics shown in Table 1 has the form (2.1)-(2.2) by the law of mass action:

$$\frac{du}{dt} = \alpha(t) - \beta u(t), \quad (2.1)$$

$$\frac{ds}{dt} = \beta u(t) - \gamma s(t), \quad (2.2)$$

where $t \geq 0$, $(u(t), s(t))|_{t=0} = (u_0, s_0)$, and

$$\alpha(t) = \begin{cases} \alpha^{\text{on}}, & t \leq t_s, \\ \alpha^{\text{off}} = 0, & t > t_s. \end{cases} \quad (2.3)$$

Here t_s is the switch time of the transcriptional process.

The term defined through Eq. (2.2):

$$\mathbf{v}(t) = (v_g(t))_g = \left(\frac{ds_g}{dt} \right)_g = (\beta_g u_g(t) - \gamma_g s_g(t))_g \in \mathbb{R}^{n_g} \quad (2.4)$$

is the RNA velocity of each cell, where $g = 1 : n_g$ in Eq. (2.4) and n_g is the number of considered genes in the RNA-seq data. Note that the velocity \mathbf{v} only depends on the state (u, s) , but not the absolute magnitude of t , given the rate parameters, since the considered system is autonomous.

Explicit solution. We will study the cases $\beta \neq \gamma$ and $\beta = \gamma$, respectively.

Case 1: $\beta \neq \gamma$. The analytical solution to (2.1)-(2.2) in the *on* stage with rate $\alpha^{\text{on}} = \alpha$ is

$$u(t) = u_0 e^{-\beta t} + \frac{\alpha}{\beta} (1 - e^{-\beta t}), \tag{2.5}$$

$$s(t) = s_0 e^{-\gamma t} + \frac{\alpha}{\gamma} (1 - e^{-\gamma t}) + \frac{\alpha - \beta u_0}{\gamma - \beta} (e^{-\gamma t} - e^{-\beta t}) \tag{2.6}$$

for $t \leq t_s$. Usually, we suppose $(u_0, s_0) = (0, 0)$, then we have

$$u(t) = \frac{\alpha}{\beta} (1 - e^{-\beta t}), \tag{2.7}$$

$$s(t) = \frac{\alpha}{\gamma} (1 - e^{-\gamma t}) + \frac{\alpha}{\gamma - \beta} (e^{-\gamma t} - e^{-\beta t}) \tag{2.8}$$

for $t \leq t_s$. Define the switch state by

$$(u_s, s_s) := (u(t_s), s(t_s)).$$

Then in the *off* stage, we have the solution

$$u(t) = u_s e^{-\beta(t-t_s)}, \tag{2.9}$$

$$s(t) = s_s e^{-\gamma(t-t_s)} - \frac{\beta u_s}{\gamma - \beta} (e^{-\gamma(t-t_s)} - e^{-\beta(t-t_s)}) \tag{2.10}$$

for $t > t_s$. It is straightforward that $u(t), s(t) > 0$ for any finite $t > 0$.

Case 2: $\beta = \gamma$. The analytical solution to (2.1)-(2.2) in the *on* stage is

$$u(t) = u_0 e^{-\beta t} + \frac{\alpha}{\beta} (1 - e^{-\beta t}), \tag{2.11}$$

$$s(t) = s_0 e^{-\beta t} + \frac{\alpha}{\beta} (1 - e^{-\beta t}) - (\alpha - \beta u_0) t e^{-\beta t} \tag{2.12}$$

for $t \leq t_s$. When $(u_0, s_0) = (0, 0)$, we have

$$u(t) = \frac{\alpha}{\beta} (1 - e^{-\beta t}), \tag{2.13}$$

$$s(t) = \frac{\alpha}{\beta} (1 - e^{-\beta t}) - \alpha t e^{-\beta t} \tag{2.14}$$

for $t \leq t_s$. And in the *off* stage

$$u(t) = u_s e^{-\beta(t-t_s)}, \tag{2.15}$$

$$s(t) = s_s e^{-\beta(t-t_s)} + u_s \beta (t - t_s) e^{-\beta(t-t_s)} \tag{2.16}$$

for $t > t_s$. It is straightforward to note that the solution (2.11)-(2.12) is indeed the limit of the solution (2.5)-(2.6) as $\gamma \rightarrow \beta$.

Steady State. The steady state in the on stage is $(u^*, s^*) = \alpha / (\beta, \gamma)$, and the steady state in the off stage is simply $(u_*, s_*) = (0, 0)$.

Scale Invariance. It is important to note that the system (2.1)-(2.2) has the following scale invariance property, i.e. if we define the parameter $\theta = (\theta_r, t_s)$, where the rates $\theta_r = (\alpha, \beta, \gamma)$, then the solution satisfies

$$(u(t; \theta_r, t_s), s(t; \theta_r, t_s)) = (u(\kappa t; \theta_r / \kappa, \kappa t_s), s(\kappa t; \theta_r / \kappa, \kappa t_s)) \quad (2.17)$$

for any scaling parameter $\kappa > 0$. This scale invariance indicates the degeneracy of the inference problem. That is, to ensure the well-posedness of the inference on parameter θ , we should fix the time scale of the system. For example, one can consider the dynamics (2.1)-(2.2) within a fixed period $[0, T]$, where $t_{\max} = T$. We remark that the choice of the degree of freedom does affect the magnitude of RNA velocity (2.4) up to a multiplicative constant.

2.2 Stochastic model

In the stochastic model, the system state $(u(t), s(t)) \in \mathbb{N}^2$ is a stochastic process and we are interested in the evolution of its probability mass function denoted by

$$P_{mn}(t) := \text{Prob} \{ (u(t), s(t)) = (m, n) \in \mathbb{N}^2 \}. \quad (2.18)$$

The stochastic model of the reaction dynamics shown in Table 1 is given by the following chemical master equation (CME) [19]

$$\begin{aligned} \partial_t P_{mn} = & \alpha^{\text{on/off}} (P_{m-1, n} - P_{mn}) + \beta [(m+1)P_{m+1, n-1} - mP_{mn}] \\ & + \gamma [(n+1)P_{m, n+1} - nP_{mn}] \end{aligned} \quad (2.19)$$

with initial condition $P_{mn}(0) = p_{mn}^0$. In the on or off stage, the production rate of u , $\alpha^{\text{on/off}}$, will be set as $\alpha^{\text{on}} = \alpha$ or $\alpha^{\text{off}} = 0$, respectively. We will study the analytical solution of (2.19) in different cases.

Scale Invariance. Similar to the deterministic case, the solution $P_{mn}(t)$ has the scale invariance property

$$P_{mn}(t; \theta_r, t_s) = P_{mn}(\kappa t; \theta_r / \kappa, \kappa t_s) \quad (2.20)$$

for any scaling parameter $\kappa > 0$.

2.2.1 On stage with zero initial value

In the on stage, i.e. $t \leq t_s$, we have the CME

$$\begin{aligned} \partial_t P_{mn} = & \alpha (P_{m-1, n} - P_{mn}) + \beta [(m+1)P_{m+1, n-1} - mP_{mn}] \\ & + \gamma [(n+1)P_{m, n+1} - nP_{mn}]. \end{aligned} \quad (2.21)$$

We will first study the case $(u(0),s(0))=(0,0)$, i.e. with initial distribution $P_{mn}(0)=\delta_{m0}\delta_{n0}$, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i=j, \\ 0 & \text{otherwise} \end{cases}$$

is the Kronecker's delta-function. The general cases are left in Section 2.2.3. Since the rate functions are all linear in m and n , we will employ the idea of moment generating function to solve (2.21) [45].

Theorem 2.1 (Analytical Distribution in the On Stage). *With initial distribution $P_{mn}(0) = \delta_{m0}\delta_{n0}$, the solution of Eq. (2.21) is*

$$P_{mn|00}^{\text{on}}(t) = \frac{a^m(t)b^n(t)}{m!n!} e^{-a(t)-b(t)}, \quad (m,n) \in \mathbb{N}^2, \tag{2.22}$$

where

$$a(t) = \frac{\alpha}{\beta}(1 - e^{-\beta t}), \tag{2.23}$$

$$b(t) = \begin{cases} \frac{\alpha}{\gamma}(1 - e^{-\gamma t}) + \frac{\alpha}{\gamma - \beta}(e^{-\gamma t} - e^{-\beta t}), & \beta \neq \gamma, \\ \frac{\alpha}{\beta}(1 - e^{-\beta t}) - \alpha t e^{-\beta t}, & \beta = \gamma, \end{cases} \tag{2.24}$$

and the notation $P_{mn|00}^{\text{on}}(t)$ stands for the transition probability from state $(0,0)$ to (m,n) .

Proof. Consider the moment generating function $F(y,z,t) = \sum_{m,n} y^m z^n P_{m,n}(t)$. Then we have

$$\partial_t F = \alpha(yF - F) + \beta(z\partial_y F - y\partial_y F) + \gamma(\partial_z F - z\partial_z F) \tag{2.25}$$

with initial condition $F_0(y,z) = \sum_{m,n} y^m z^n P_{mn}(0)$. Under the zero initial value condition on $(u(t),s(t))$, we have $F_0 = 1$.

Introduce the change of variable

$$u = y - 1, \quad v = z - 1, \quad \tilde{F}(u,v,t) = F(u+1,v+1,t) = F(y,z,t).$$

Then from (2.25) we get

$$\partial_v \tilde{F} + \frac{\beta}{\gamma} \left(\frac{u}{v} - 1 \right) \partial_u \tilde{F} + \frac{1}{\gamma v} \partial_t \tilde{F} = \frac{\alpha u}{\gamma v} \tilde{F}. \tag{2.26}$$

By the method of characteristics, we introduce the auxiliary variable r :

$$\frac{dv}{dr} = 1, \quad \frac{dt}{dr} = \frac{1}{\gamma v}, \quad \frac{du}{dr} = \frac{\beta}{\gamma} \left(\frac{u}{v} - 1 \right), \quad \frac{d\tilde{F}}{dr} = \frac{\alpha u}{\gamma v} \tilde{F} \tag{2.27}$$

with initial condition $(u, v, t, \tilde{F})|_{r=0} = (u_0, v_0, 0, \tilde{F}_0)$, where $\tilde{F}_0 = F_0(u_0 + 1, v_0 + 1)$.

Case 1: $\beta \neq \gamma$. Solving (2.27), we get $v = r + v_0$, $dv = dr$, $t = \gamma^{-1} \ln v / v_0$, and correspondingly

$$v = v_0 e^{\gamma t}. \quad (2.28)$$

For u , we obtain

$$\frac{du}{dv} = \frac{\beta u}{\gamma v} - \frac{\beta}{\gamma}, \quad \text{which induces } u = \left(u_0 - \frac{\beta v_0}{\beta - \gamma} \right) \left(\frac{v}{v_0} \right)^{\frac{\beta}{\gamma}} + \frac{\beta}{\beta - \gamma} v. \quad (2.29)$$

Define $a_0 := u_0 - (\beta - \gamma)^{-1} \beta v_0$, we get

$$\frac{d\tilde{F}}{dv} = \frac{\alpha u}{\gamma v} \tilde{F} = \frac{\alpha}{\gamma} \left(a_0 v_0^{-\frac{\beta}{\gamma}} v^{\frac{\beta}{\gamma} - 1} + \frac{\beta}{\beta - \gamma} \right) \tilde{F},$$

thus

$$\ln \frac{\tilde{F}}{\tilde{F}_0} = \frac{\alpha}{\beta} a_0 \left(\frac{v}{v_0} \right)^{\frac{\beta}{\gamma}} + \frac{\alpha \beta}{\gamma(\beta - \gamma)} (v - v_0) - \frac{\alpha}{\beta} a_0. \quad (2.30)$$

With initial distribution $P_{m,n}(0) = \delta_{m0} \delta_{n0}$, we have $\tilde{F}_0 = 1$. Combining Eqs. (2.28), (2.29), (2.30) and the definition of a_0 , we get

$$\tilde{F}(u, v, t) = \exp \left\{ \frac{\alpha}{\beta} (1 - e^{-\beta t}) u + \frac{\alpha \beta v}{\beta - \gamma} \left[\frac{1}{\gamma} (1 - e^{-\gamma t}) - \frac{1}{\beta} (1 - e^{-\beta t}) \right] \right\}. \quad (2.31)$$

After suitable manipulation, we obtain

$$F(y, z, t) = e^{ya(t)} \cdot e^{zb(t)} \cdot e^{-a(t) - b(t)}, \quad (2.32)$$

where $a(t)$ and $b(t)$ are defined in (2.23) and (2.24) when $\beta \neq \gamma$, respectively.

Case 2: $\beta = \gamma$. We can show that (2.29) will be replaced by

$$\frac{du}{dv} = \frac{u}{v} - 1, \quad \text{thus } u = v \left(\frac{u_0}{v_0} - \ln \frac{v}{v_0} \right). \quad (2.33)$$

After suitable derivations, we get

$$\tilde{F}(u, v, t) = \exp \left\{ \frac{\alpha}{\beta} (1 - e^{-\beta t}) (u + v) - \alpha t v e^{-\beta t} \right\}, \quad (2.34)$$

and finally we have the same formula (2.32), while $a(t)$ and $b(t)$ are defined in (2.23) and (2.24) when $\beta = \gamma$, respectively.

Therefore,

$$P_{m,n}(t) = \frac{1}{m!n!} \frac{\partial^{m+n}}{\partial y^m \partial z^n} F \Big|_{(y,z)=(0,0)} = \frac{a^m(t) b^n(t)}{m!n!} e^{-a(t) - b(t)}.$$

In summary, $u(t)$ and $s(t)$ are independently Poisson distributed with mean $a(t)$ and $b(t)$, respectively. \square

Remark 2.1. It is not surprising to observe that the mean $a(t), b(t)$ are exactly the abundance of u and s in Eqs. (2.7)-(2.8) or (2.13)-(2.14) in the deterministic model, which is well-known due to the linearity of the rates. However, Theorem 2.1 further states that u and s are independently Poisson distributed, which is not a straightforward result.

Invariant Distribution. It is obvious that the invariant distribution of (u, s) in this case is independent Poisson with parameters (u^*, s^*) .

2.2.2 Off stage with general initial data

We will study the off stage case in this section. Now we have $\alpha^{\text{off}} = 0$ and the CME is

$$\partial_t P_{mn} = \beta[(m+1)P_{m+1, n-1} - mP_{mn}] + \gamma[(n+1)P_{m, n+1} - nP_{mn}]. \tag{2.35}$$

We first consider the case with initial value $(u(0), s(0)) = (M, N)$.

Theorem 2.2 (Analytical Distribution in the Off Stage). *Define*

$$p_1(t) = e^{-\beta t}, \tag{2.36}$$

$$p_2(t) = \begin{cases} \frac{\beta}{\beta - \gamma} \frac{e^{-\gamma t} - e^{-\beta t}}{1 - e^{-\beta t}}, & \beta \neq \gamma, \\ \beta t \frac{e^{-\beta t}}{1 - e^{-\beta t}}, & \beta = \gamma, \end{cases} \tag{2.37}$$

$$p_3(t) = e^{-\gamma t}, \tag{2.38}$$

and $q_i(t) = 1 - p_i(t)$ correspondingly. We have $p_i(t), q_i(t) \in [0, 1]$ for $i = 1, 2, 3$. Then, with initial distribution $P_{mn}(0) = \delta_{mM} \delta_{nN}$, the solution of Eq. (2.35) has the form

$$P_{mn|MN}^{\text{off}}(t) = P_{m|M}^{\text{off}}(t) \cdot P_{n|mN}^{\text{off}}(t) = B_m(M, p_1) \cdot C_n(M - m, p_2, N, p_3), \tag{2.39}$$

where $m \leq M, n \leq N + M - m$,

$$B_m(M, p_1) = \binom{M}{m} p_1^m (1 - p_1)^{M - m} \tag{2.40}$$

is the probability of the binomial distribution $\mathcal{B}(M, p_1(t))$, and

$$C_n(M - m, p_2, N, p_3) = \sum_{k=0}^n B_k(M - m, p_2) B_{n-k}(N, p_3) \tag{2.41}$$

is the probability of the sum of two independent binomials $\mathcal{B}(M - m, p_2(t))$ and $\mathcal{B}(N, p_3(t))$. We take the convention that $B_k(M - m, p_2) = 0$ if $k > M - m$.

Proof. First let us show that $p_2(t) \in [0, 1]$. When $\beta \neq \gamma$, we have

$$p_2(t) = \frac{f((\beta - \gamma)t)}{f(\beta t)}, \quad \text{where } f(x) = \frac{e^x - 1}{x}.$$

We have $f'(x) = (e^x(x-1) + 1)/x^2 \geq 0$ since $g(0) = 0$, and $g'(x) > 0$ for $x > 0$ and $g'(x) < 0$ for $x < 0$, where $g(x) := e^x(x-1) + 1$. The case $\beta = \gamma$ is trivial by observing that the function $x/(e^x - 1) \in [0, 1]$ for $x \geq 0$.

Next we derive the distribution P_{mn} . Similar to the proof of Theorem 2.1, for moment generating function $F(y, z, t) = \sum_{m,n} y^m z^n P_{m,n}(t)$, we have

$$\partial_t F = \beta(z\partial_y F - y\partial_y F) + \gamma(\partial_z F - z\partial_z F).$$

Similarly define

$$u = y - 1, \quad v = z - 1, \quad \tilde{F}(u, v, t) = F(u + 1, v + 1, t) = F(y, z, t),$$

we obtain

$$\partial_v \tilde{F} + \frac{\beta}{\gamma} \left(\frac{u}{v} - 1 \right) \partial_u \tilde{F} + \frac{1}{\gamma v} \partial_t \tilde{F} = 0.$$

Introduce the parameter r , we get by the method of characteristics

$$\frac{dv}{dr} = 1, \quad \frac{dt}{dr} = \frac{1}{\gamma v}, \quad \frac{du}{dr} = \frac{\beta}{\gamma} \left(\frac{u}{v} - 1 \right), \quad \frac{d\tilde{F}}{dr} = 0. \quad (2.42)$$

Case 1: $\beta \neq \gamma$. Similar derivation shows

$$v_0 = e^{-\gamma t} v, \quad u_0 = e^{-\beta t} \left(u - \frac{\beta}{\beta - \gamma} v \right) + \frac{\beta}{\beta - \gamma} e^{-\gamma t} v.$$

We obtain

$$\begin{aligned} \tilde{F}(u, v, t) &= \tilde{F}_0 = (u_0 + 1)^M (v_0 + 1)^N \\ &= \left[e^{-\beta t} \left(u - \frac{\beta}{\beta - \gamma} v \right) + \frac{\beta}{\beta - \gamma} e^{-\gamma t} v + 1 \right]^M [e^{-\gamma t} v + 1]^N. \end{aligned}$$

After suitable manipulations, we get

$$\begin{aligned} F(y, z, t) &= \left(p_1(t)y + q_1(t) \cdot (p_2(t)z + q_2(t)) \right)^M \left(p_3(t)z + q_3(t) \right)^N \\ &= \sum_{k=0}^M \binom{M}{k} p_1^k(t) y^k q_1^{M-k}(t) \cdot (p_2(t)z + q_2(t))^{M-k} (p_3(t)z + q_3(t))^N, \end{aligned} \quad (2.43)$$

which exactly has the probabilistic interpretation as shown in Eqs. (2.40)-(2.41).

Case 2: $\beta = \gamma$. We can show that

$$v_0 = e^{-\gamma t} v, \quad u_0 = e^{-\beta t} u + \beta t e^{-\beta t} v$$

in this case. Substitute into $\tilde{F} = (u_0 + 1)^M (v_0 + 1)^N$, we get the same equation (2.43) but with the $p_i(t)$ in the $\beta = \gamma$ case. \square

Remark 2.2. In the off stage, the unspliced mRNA u obeys the binomial distribution. Given $u(t) = m$, the conditional distribution of the spliced mRNA s is characterized by the sum of two independent binomial random variables. Intuitively, s is comprised of two parts: the new spliced mRNA counts generated from u and the non-degraded spliced mRNA counts from the initial state. It is also natural to observe that the mean of $(u(t), s(t))$ is $(Mp_1(t), Mq_1(t)p_2(t) + Np_3(t))$, which is essentially (2.9)-(2.10).

Corollary 2.1. When the initial distribution $P_{mn}(0) = p_{mn}^0$, the solution of Eq. (2.35) is

$$P_{mn}^{\text{off}}(t) = \sum_{k \geq m, l \geq n} P_{mn|kl}^{\text{off}}(t) p_{kl}^0 \tag{2.44}$$

for $(m, n) \in \mathbb{N}^2$.

Invariant Distribution. It is straightforward that the invariant distribution in the off stage is simply $P_{mn}(\infty) = \delta_{m0} \delta_{n0}$ as $t \rightarrow \infty$.

2.2.3 On stage with general initial data

Corollary 2.2. When the initial distribution $P_{mn}(0) = \delta_{mm_0} \delta_{nn_0}$, the solution of Eq. (2.21) is

$$P_{mn|m_0n_0}^{\text{on}}(t) = \sum_{k=0}^m \sum_{l=0}^n P_{kl|00}^{\text{on}}(t) P_{m-k, n-l|m_0n_0}^{\text{off}}(t), \tag{2.45}$$

which is the convolution of the distributions $P_{mn|00}^{\text{on}}(t)$ and $P_{mn|m_0n_0}^{\text{off}}(t)$. We adopt the convention that $P_{mn|m_0n_0}^{\text{off}}(t) = 0$ if $m > m_0$ or $n > n_0$.

Proof. To check the result, we only need to note that \tilde{F}_0 in (2.30) will be replaced by $(u_0 + 1)^{m_0} (v_0 + 1)^{n_0}$. Thus

$$F(y, z, t) = F^{\text{on}}(y, z, t) F^{\text{off}}(y, z, t),$$

where $F^{\text{on}}(y, z, t)$ and $F^{\text{off}}(y, z, t)$ are defined as in (2.32) and (2.43), respectively. This naturally yields to the transition probability (2.45). \square

Remark 2.3. The above result leads to a recipe to directly generate samples from $P_{mn|m_0n_0}^{\text{on}}(t)$:

$$(u(t), v(t)) = (X_1, Y_1) + (X_2, Y_2 + Y_3),$$

where X_1, Y_1, X_2, Y_3 are independent random variables with $X_1 \sim \mathcal{P}(a(t))$, $Y_1 \sim \mathcal{P}(b(t))$, $X_2 \sim \mathcal{B}(m_0, p_1(t))$, $Y_3 \sim \mathcal{B}(n_0, p_3(t))$, and $Y_2 | X_2 \sim \mathcal{B}(m_0 - X_2, p_2(t))$.

Corollary 2.3. *When the initial distribution $P_{mn}(0) = p_{mn}^0$, the solution of Eq. (2.21) is*

$$P_{mn}^{\text{on}}(t) = \sum_{k,l \in \mathbb{N}} P_{mn|kl}^{\text{on}}(t) p_{kl}^0, \quad (m,n) \in \mathbb{N}^2. \quad (2.46)$$

3 Inference of RNA velocity

In this section, we will study the inverse problem: the inference of the parameters in the RNA velocity model from the data. We will mainly revisit the proposals pursued in [3,29] and briefly mention our new approach [60], which utilizes the full stochastic model to do the inference.

3.1 Steady state model

The steady state model was first considered in [29]. In this model, one assumes that the on stage lasts sufficiently long so the state of the system is close to the steady state of the dynamical system (2.1)-(2.2). Therefore, the upper-right corner points in the (u,s) -plot can be approximated as steady states. In the steady state, we have

$$\frac{ds}{dt} = \beta u(t) - \gamma s(t) = 0,$$

which means that the mRNA synthesis and degradation are in balance. This balance condition in the steady state can be utilized to approximate the ratio of degradation and splicing rates via least squares fitting as

$$v^* = \left(\frac{\gamma}{\beta} \right)^* = \underset{v}{\operatorname{argmin}} \beta^2 \| \mathbf{u} - v \mathbf{s} \|^2 = \frac{\mathbf{u}^T \mathbf{s}}{\| \mathbf{s} \|^2}, \quad (3.1)$$

where \mathbf{u}, \mathbf{s} are vectors with components corresponding to the cells in the upper-right corner points in the (u,s) -plot for each gene. If further assuming that $\beta = 1$ across all genes in [29] via scale invariance argument, the RNA velocity is then estimated as

$$\mathbf{v} = (v_g)_g, \quad v_g = u_g - v_g^* s_g.$$

Though original, simple and successful, the above steady state model and the treatment with $\beta = 1$ for all genes are not good enough assumptions in many cases. In fact, setting the splicing rates $\beta = 1$ for all genes is actually wrong according to the scale invariance property of the system, which only permits one degree of freedom to be adjusted. These drawbacks call for more robust and accurate estimation methods for the RNA velocity.

3.2 EM algorithm for the transient models

In this subsection, we will revisit and study the parameter inference using EM algorithm for transient models. The most related references on this aspect are [3, 60]. To ensure the computational feasibility, we also employ suitable approximations, which will be stated in corresponding places below.

3.2.1 Basic framework of EM

Given the observed data $X = (x_{cg})_{c=1:n_c, g=1:n_g}$, where $x_{cg} = (u_{cg}, s_{cg})$ for cell c and gene g , we want to maximize the log-likelihood

$$\begin{aligned} L(\theta; X) &= \ln P(X|\theta) = \ln \prod_{c=1}^{n_c} \prod_{g=1}^{n_g} P(x_{cg}|\theta_g) \\ &= \sum_{c=1}^{n_c} \sum_{g=1}^{n_g} \ln P(x_{cg}|\theta_g) := \sum_{c=1}^{n_c} \sum_{g=1}^{n_g} l(\theta_g; x_{cg}). \end{aligned} \quad (3.2)$$

Note that the abundance x_{cg} depends on time t_{cg} and the switch state $x_{s,cg} = (u_{s,cg}, s_{s,cg})$, which are not observables, we indeed encounter a hidden variable problem. It is natural to utilize the EM algorithm to do the inference.

Let us introduce the latent variable $h_{cg} = (t_{cg}, x_{s,cg})$ for $c = 1:n_c, g = 1:n_g$, where t is the latent time and x_s is the latent switch state. Then the log-likelihood function can be written as

$$\begin{aligned} l(\theta_g; x_{cg}) &= \ln P(x_{cg}|\theta_g) = \ln P(x_{cg}, h_{cg}|\theta_g) - \ln P(h_{cg}|x_{cg}; \theta_g) \\ &:= l_0(\theta_g; x_{cg}, h_{cg}) + l_1(\theta_g; h_{cg}|x_{cg}). \end{aligned} \quad (3.3)$$

Similarly denote the sum of $l_0(\theta_g; x_{cg}, h_{cg})$ and $l_1(\theta_g; h_{cg}|x_{cg})$ with respect to c, g as $L_0(\theta; X, h)$ and $L_1(\theta; h|X)$, respectively. Then

$$L(\theta; X) = L_0(\theta; X, h) + L_1(\theta; h|X).$$

Taking conditional expectation with respect to the distribution of $h|X$ given parameter θ' , we get

$$\begin{aligned} L(\theta; X) &= \mathbb{E}_{h|X, \theta'} [L_0(\theta; X, h)] + \mathbb{E}_{h|X, \theta'} [L_1(\theta; h|X)] \\ &:= Q(\theta|\theta') + R(\theta|\theta'), \end{aligned}$$

where

$$R(\theta|\theta') = - \int P(h|X; \theta') \ln P(h|X; \theta) dh.$$

The above formulation is the basis of the well-known EM algorithm [11], which can be stated as below.

Algorithm 1 EM Algorithm for the RNA Velocity Model.

1. Start with an initial guess for parameter $\theta^{(0)}$. Set $j=0$.
2. *Expectation Step*: at the $(j+1)$ th step, compute

$$Q(\theta|\theta^{(j)}) = \mathbb{E}_{h|X,\theta^{(j)}}(L_0(\theta; X, h))$$

as a function of θ .

3. *Maximization Step*: determine the new estimate $\theta^{(j+1)}$ as the maximizer of $Q(\theta|\theta^{(j)})$ over θ .
 4. Iterate Steps 2 and 3 until convergence.
-

It is a classical result that the EM iterations never decreases the log-likelihood $L(\theta; X)$. In fact, if θ maximizes $Q(\theta|\theta')$, we have

$$L(\theta; X) - L(\theta'; X) = (Q(\theta|\theta') - Q(\theta'|\theta')) + (R(\theta|\theta') - R(\theta'|\theta')) \geq 0$$

since $R(\theta|\theta') - R(\theta'|\theta') = D_{\text{KL}}(P(h|X;\theta') || P(h|X;\theta)) \geq 0$ by the non-negativity of Kullback-Leibler divergence [10]. This feature guarantees the local convergence of EM iterations.

3.2.2 EM for the deterministic model

For the deterministic RNA velocity model, the latent variable h can be reduced to t since the switch state $x_s = x(t_s; \theta)$ is uniquely determined. So we will replace h with t in (3.3) in the deterministic setup. If we assume the observation noise is Gaussian with mean 0 and variance σ^2 for all cells and genes, and the sampling time t_{cg} is uniformly distributed in a fixed period $[0, T]$, we have

$$\begin{aligned} P(x_{cg}, t_{cg} | \theta_g) &= P(x_{cg} | t_{cg}; \theta_g) \cdot P(t_{cg} | \theta_g) \\ &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|x_{cg} - x(t_{cg}; \theta_g)|^2}{2\sigma^2}\right) \cdot \frac{1}{T}, \end{aligned} \quad (3.4)$$

where $x(t_{cg}; \theta_g)$ is the solution of (2.1)-(2.2) at time t_{cg} with parameter θ_g . Then

$$l_0(\theta_g; x_{cg}, t_{cg}) = \ln P(x_{cg}, t_{cg} | \theta_g) \propto -|x_{cg} - x(t_{cg}; \theta_g)|^2,$$

and

$$P(t_{cg} | x_{cg}; \theta_g) = \frac{P(x_{cg}, t_{cg} | \theta_g)}{P(x_{cg} | \theta_g)} \propto \exp\left(-\frac{|x_{cg} - x(t_{cg}; \theta_g)|^2}{2\sigma^2}\right), \quad (3.5)$$

where $P(x_{cg}|\theta_g)$ is ignored for its independence with t_{cg} . So we obtain

$$\begin{aligned} L_0(\theta; X, t) &\propto -\|X - X(t; \theta)\|^2, \\ P(t|X; \theta) &\propto \exp\left(-\frac{\|X - X(t; \theta)\|^2}{2\sigma^2}\right), \end{aligned}$$

where $X(t, \theta) := (x(t_{cg}; \theta_g))_{c=1:n_c; g=1:n_g}$.

According to EM Algorithm, we have

$$\begin{aligned} \theta^{(j+1)} &= \operatorname{argmax}_{\theta} \int_0^T L_0(\theta; X, t) \cdot P(t|X, \theta^{(j)}) dt \\ &= \operatorname{argmin}_{\theta} \int_0^T \|X - X(t; \theta)\|^2 \exp\left(-\frac{\|X - X(t; \theta^{(j)})\|^2}{2\sigma^2}\right) dt. \end{aligned} \quad (3.6)$$

In the small noise limit regime, i.e. $\sigma \rightarrow 0$, by Laplace asymptotics [2], we get

$$\theta^{(j+1)} = \operatorname{argmin}_{\theta} \|X - X(t^{(j)}; \theta)\|^2, \quad (3.7)$$

while

$$t^{(j)} = \operatorname{argmin}_t \|X - X(t; \theta^{(j)})\|^2. \quad (3.8)$$

It forms an iteration between the parameter θ and the latent time t . Below we discuss more detailed procedure in (3.7)-(3.8).

Update of t . Two different models can be utilized in the update step (3.8), which we term the *independent- t model* and *uniform- t model* below. The two different choices lead to different computational complexity. We further assume that the switch time $t_s = (t_{s,g})_{g=1:n_g}$ is only gene dependent throughout the transient model estimations.

Independent- t model. In this model, we permit the time t_{cg} for different g to be different, i.e. for a specific cell c , $(t_{cg})_{g=1:n_g}$ are independent. So we can estimate t_{cg} for each c and g separately.

The fact, that the estimation of $(t_{cg})_{c=1:n_c}$ for different g can be separated, tells that we only need to consider a fixed g . Given θ_g and $(x_{cg})_{1:n_c}$, to estimate the optimal $(t_{cg})_{1:n_c}$, we classify the state of cell c into the on state if $t_{cg} \leq t_{s,g}$ or off state otherwise. Define the objective function

$$d(t_{cg}; x_{cg}, \theta_g) = |x_{cg} - x(t_{cg}; \theta_g)|^2, \quad (3.9)$$

which has only two piecewise smooth parts determined by $t_{s,g}$. We first compute the optimal t_{cg}^{on} and t_{cg}^{off} by assuming the cell c is in on or off stage, respectively with

$$t_{cg}^{\text{on}} = \operatorname{argmin}_{t_{cg} \leq t_{s,g}} d(t_{cg}; x_{cg}, \theta_g), \quad t_{cg}^{\text{off}} = \operatorname{argmin}_{t_{cg} \geq t_{s,g}} d(t_{cg}; x_{cg}, \theta_g).$$

This optimization is feasible since we know the analytical form of $x(t_{cg};\theta_g)$. The optimal estimation of t_{cg} is then obtained by

$$t_{cg} = \begin{cases} t_{cg}^{\text{on}}, & \text{if } d(t_{cg}^{\text{on}};x_{cg},\theta_g) < d(t_{cg}^{\text{off}};x_{cg},\theta_g), \\ t_{cg}^{\text{off}}, & \text{otherwise.} \end{cases}$$

The computational complexity in this setup is $\mathcal{O}(2n_g)$ in terms of the analytical function evaluations of $x(t;\theta)$.

Uniform- t model. In this model, we require that the time t_{cg} for different genes are consistent, i.e. $t_{cg} = t_c$ for $g = 1:n_g$. This setup is more reasonable in reality, however, it brings difficulty into the optimization.

In the uniform- t model, the objective functions (3.9) are no longer separated for different genes. Instead, we should consider the minimization of

$$d(t_c;x_c,\theta) = \sum_g |x_{cg} - x(t_c;\theta_g)|^2, \quad (3.10)$$

which generally has $n_g + 1$ piecewise smooth parts determined by $(t_{s,g})_{g=1:n_g}$. We can first sort the switch time t_s like

$$t_{s,g_1} \leq t_{s,g_2} \leq \dots \leq t_{s,g_{n_g}},$$

and next compute t_c^k by minimizing (3.10) in each subinterval $[t_{s,g_k}, t_{s,g_{k+1}})$ for $k=0,1,\dots,n_g$ with $t_{s,g_0} = 0$ and $t_{s,g_{n_g+1}} = T$. Finally we can obtain an optimal t_c which minimizes the objective functions $d(t_c^k;x_c,\theta)$.

The computational complexity in this setup is $\mathcal{O}(n_g(n_g + 1))$ compared with the independent t model.

Update of θ . For different genes g_1 and g_2 , the update of θ_{g_1} and θ_{g_2} are independent according to (3.7). So we fix g and consider $\theta_g = (\alpha_g, \beta_g, \gamma_g, t_{s,g})$. There is no difference on the update of θ_g with independent- t model or uniform- t model because of the independence. We will only use the independent- t model formulation for illustration in the following text.

Define the objective function

$$d(\theta_g;x_g,t_g) = \sum_c |x_{cg} - x(t_{cg};\theta_g)|^2. \quad (3.11)$$

The minimization of (3.11) can be performed in different ways. One possible approach is to do the optimization with respect to the rates $(\alpha_g, \beta_g, \gamma_g)$ and the switch time $t_{s,g}$ alternatively, due to the non-smoothness induced by $t_{s,g}$. Another approach is to reduce (3.11) into a new function of rates only

$$\tilde{d}(\alpha_g, \beta_g, \gamma_g) = \min_{t_{s,g}} d(\theta_g;x_g,t_g),$$

and then optimize \tilde{d} directly. In any case, the objective function is a sum of n_c terms with non-smoothness induced by $t_{s,g}$. The optimization is not easy and usually stuck in the local minimum with local convergence methods.

Remark 3.1. We remark here that the choice of $P(t_{cg}|\theta_g) = 1/T$ can be altered, e.g. the uniform distribution along the curve $(u(t),s(t))_{t \in [0,T]}$ or other proposals, However, this will increase the computational complexity. When other choices of the observation noise are taken, the relative scales of the noise on different genes/cells still remain in the norm $\|\cdot\|$ in (3.7)-(3.8) even if we consider the zero noise limit regime.

The above EM algorithm with independent- t model in the zero noise limit regime is utilized in [3].

3.2.3 EM for the stochastic model

It is natural to consider the inference of the stochastic RNA velocity model. In the stochastic setup, the randomness of the switch state x_s should be incorporated into the full likelihood. Similarly assume that the observation noise is Gaussian with mean 0 and variance σ^2 for all cells and genes, and the uniform distribution on the sampling time t_{cg} in a fixed period $[0, T]$. We have

$$P(x_{cg}, h_{cg} | \theta_g) = P(x_{cg} | h_{cg}; \theta_g) P(x_{s,cg} | \theta_g) P(t_{cg} | \theta_g),$$

where

$$P(t_{cg} | \theta_g) = \frac{1}{T}, \quad P(x_{s,cg} | \theta_g) = P_{x_{s,cg}|0}^{\text{on}}(t_{s,g}; \theta_g),$$

$$P(x_{cg} | h_{cg}; \theta_g) = \sum_y P_y(t_{cg}; \theta_g) \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|y - x_{cg}|^2}{2\sigma^2}\right),$$

and

$$P_y(t_{cg}; \theta_g) = \begin{cases} P_{y|0}^{\text{on}}(t_{cg}; \theta_g), & t_{cg} \leq t_{s,g}, \\ P_{y|x_{s,cg}}^{\text{off}}(t_{cg}; \theta_g), & t_{cg} > t_{s,g}. \end{cases} \quad (3.12)$$

In the zero noise limit $\sigma \rightarrow 0$, we get

$$l_0(x_{cg}, h_{cg} | \theta_g) \propto \ln P_{[x_{cg}]}(t_{cg}; \theta_g) + \ln P_{x_{s,cg}|0}^{\text{on}}(t_{s,g}; \theta_g)$$

in the leading order, where $[x]$ is the Gaussian nearest integer function. So we have

$$L_0(X, h | \theta) = \sum_{c,g} l_0(x_{cg}, h_{cg} | \theta_g)$$

and correspondingly

$$\theta^{(j+1)} = \underset{\theta}{\operatorname{argmax}} \mathbb{E}_h |_{X, \theta^{(j)}} L_0(X, h | \theta).$$

The optimization of the above formulation is not trivial. We leave further algorithmic constructions and practical applications to our continued publication [60].

4 Dynamical analysis based on RNA velocity

After estimating the rates $\theta_r = (\alpha_g, \beta_g, \gamma_g)_g$ for each gene, one can then compute the RNA velocity $v_c \in \mathbb{R}^{n_g}$ for each cell $c = 1:n_c$ via (2.4). One key component of the downstream analysis is to identify the source (stem cell) and sinks (differentiated cells) in the developmental process based on the obtained RNA velocities. We will discuss the related mathematics behind different proposals and give the rationale for this step. The continuum limit for various velocity kernels and the route-finding algorithm based on transition path theory will also be discussed.

4.1 Dynamical system view of cell-fate development

Given a single cell, let $x(t) \in \mathbb{R}^{n_g}$ denote its gene expression profile (more generally, its *state* in cell-fate development) at time t . The evolution of $x(t)$ can be described by a dynamical system model using the ordinary differential equation (ODE)

$$\frac{dx}{dt} = f(x), \quad (4.1)$$

where the vector field $f: \mathbb{R}^{n_g} \rightarrow \mathbb{R}^{n_g}$ is determined by the gene regulatory kinetics.

Since the gene expression process is subject to both extrinsic and intrinsic noise [49], it is also common to model the cell-fate transition dynamics with stochastic ordinary differential equations (SDE)

$$dx_t = f(x_t)dt + \sigma(x_t)dw_t, \quad (4.2)$$

where $w_t \in \mathbb{R}^k$ denotes the standard Wiener process, representing the noise from k reaction channels or fluctuating sources, and the matrix $\sigma \in \mathbb{R}^{n_g \times k}$ corresponds to the noise strength. The model is well-known as the chemical Langevin equation [20] with the appropriate coupling of f and σ .

As noted in [41], the RNA velocity for single cells can be incorporated in such dynamical system viewpoint to study the underlying cell-fate dynamics. Currently, there are two lines of approaches to define the dynamical system utilizing RNA velocity in existing literatures:

1. The *continuous* dynamics approach [41]. In this approach, one fits a vector field $\hat{f}(s)$ defined on the continuous space, such that $\hat{f}(s_i) \approx v_i$ for each single cell, where v_i are the estimated RNA velocities. Based on the inferred vector field \hat{f} , one can investigate the long-term dynamical properties of (4.1) or (4.2) to model the cell-fate development. The relevant important concepts include:
 - Meta-stable states of cell fates development [23], corresponding to the attractors x^* of (4.1) such that $\hat{f}(x^*) = 0$.

- Energy landscape of cell fates development [67], which is the realization of Waddington’s metaphor [55]. Given the stochastic dynamics (4.2), among the different proposals for constructing energy landscape, one popular choice is the potential landscape [56] defined as follows. Note that evolution of probability density $p(\mathbf{x}, t)$ in (4.2) can be described by the Fokker-Planck equation $\partial_t p(\mathbf{x}, t) = \mathcal{L}^* p$, where \mathcal{L}^* denotes the conjugate operator of the infinitesimal generator \mathcal{L} of SDE (4.2) such that

$$\mathcal{L}u := \hat{f} \cdot \nabla_x u + \frac{1}{2} D : \nabla_x^2 u, \quad \mathcal{L}^* u := -\nabla_x \cdot (\hat{f} u) + \frac{1}{2} \nabla_x^2 : (Du),$$

where $D(\mathbf{x}) = \sigma \sigma^T$ and $A : B$ denotes $\sum_{i,j} A_{ij} B_{ij}$. The steady-state probability distribution $p^{ss}(\mathbf{x})$ satisfies $\mathcal{L}^* p^{ss} = 0$, and the potential landscape of the system can be defined as $\phi(\mathbf{x}) = -\ln p^{ss}(\mathbf{x})$.

2. The *discrete* dynamics approach [3, 29]. With this popular proposal, one utilizes the RNA velocity v_i to construct a Markov Chain defined on individual cells, with transition probabilities $P(\mathbf{s}_i, \mathbf{s}_j)$ satisfying $\sum_{j=1}^{n_c} P(\mathbf{s}_i, \mathbf{s}_j) = 1$. Ideally, the constructed Markov Chain should approximate the continuous dynamics (4.1) or (4.2) in discrete sense. Typical applications of such discrete dynamics include calculating the steady-state distribution of the system, and detecting roots and ending cells during development.

A fundamental theoretical question regarding the two approaches, which is also of noticeable mathematical interest, lies in the *consistency* issue between the two proposals. To be more specific, could the constructed discrete dynamics (with transition probabilities defined on individual data points) *converge* into the correct continuous dynamical systems, under appropriate limit regimes? In the next subsection, we will conduct a rigorous analysis on the different continuum limits of Markov chain dynamics, which are constructed with various choices of velocity kernels in defining transition probabilities.

4.2 Defining transition probabilities among cells

Let $S = (s_{ij})_{i=1:n_c; j=1:n_g} \in \mathbb{R}^{n_c \times n_g}$ be the gene expression matrix of the spliced mRNA. We also denote

$$S = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{n_c})^T,$$

where $\mathbf{s}_i \in \mathbb{R}^{n_g}$ for $i=1:n_c$. For ease of notation, we also use $d=n_g$ for short in the following analysis.

To define the transition probability among different cells, one should take into account the randomness introduced by the unknown facts and the directed transition associated with the RNA velocity [3, 29]. The transition between two cells usually involves drift and diffusion. For the diffusion part, we consider the following two different *diffusion kernels*:

- Gaussian kernel

$$d_\epsilon(\mathbf{s}_i, \mathbf{s}_j) = h\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\epsilon}\right), \quad (4.3)$$

- kNN (k-Nearest Neighbor) kernel

$$d_n(\mathbf{s}_i, \mathbf{s}_j) = I\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\epsilon_n r_n^2(\mathbf{s}_i)}\right). \quad (4.4)$$

Here $h(\cdot)$ in (4.3) is a function with exponential decay, say $h(x) \sim \exp(-x)$ as $x \rightarrow \infty$, $I(\cdot)$ in (4.4) is an indicator function with $I(r) = 1$ for $|r| \leq 1$ and 0 otherwise, and $\sqrt{\epsilon_n} r_n(\mathbf{s}_i)$ is the location-dependent distance to the k_n th nearest neighbor of \mathbf{s}_i given n sample points ($n = n_c$ in the above setup).

The kNN kernel can be reformulated as similar form in (4.3). Following [52], we have the kNN density estimate

$$q_n(\mathbf{x}) = \frac{k_n/n}{V_d \cdot (\sqrt{\epsilon_n} r_n(\mathbf{x}))^d}, \quad (4.5)$$

where V_d is the volume of the d -dimensional unit ball and $q_n(\mathbf{x})$ will uniformly converge to the true sampling density $q(\mathbf{x})$ under mild conditions on k_n and q [12]. It is natural to choose

$$\epsilon_n = \left(\frac{k_n}{nV_d}\right)^{\frac{2}{d}} \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

and Theorems A.1 and A.2 in Appendix show that $r_n(\mathbf{x}) \xrightarrow{a.s.} r(\mathbf{x}) + o(\sqrt{\epsilon_n})$, where $r(\mathbf{x}) = (q(\mathbf{x}))^{-1/d}$. So we can still denote the kNN kernel as

$$d_\epsilon(\mathbf{s}_i, \mathbf{s}_j) = I\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\epsilon r_\epsilon^2(\mathbf{s}_i)}\right), \quad (4.6)$$

where $\epsilon = \epsilon_n \rightarrow 0$ and $r_\epsilon(\mathbf{x}) = r(\mathbf{x}) + o(\sqrt{\epsilon})$.

For the drift part, we will call it the *velocity kernel*. Bearing in mind the intuition that cell i is expected to have high probability of transition towards cell j , when the corresponding change in gene expression $\delta_{ij} = \mathbf{s}_j - \mathbf{s}_i$ matches the predicted change according to the velocity vector \mathbf{v}_i , we consider the following three schemes of velocity kernels:

- Cosine scheme

$$v(\mathbf{s}_i, \mathbf{s}_j) = g(\cos\langle \delta_{ij}, \mathbf{v}_i \rangle), \quad (4.7)$$

- Correlation scheme

$$v(\mathbf{s}_i, \mathbf{s}_j) = g(\text{corr}(\delta_{ij}, \mathbf{v}_i)), \quad (4.8)$$

- Inner-Product scheme

$$v(\mathbf{s}_i, \mathbf{s}_j) = g(\delta_{ij}^T \mathbf{v}_i), \quad (4.9)$$

where $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\text{corr}(\mathbf{x}, \mathbf{y})$ are the angle and the Pearson coefficient between the vectors \mathbf{x} and \mathbf{y} respectively, and $g(x)$ is a bounded, positive, and non-decreasing function. The overall transition kernel is then defined by

$$k_\epsilon(\mathbf{s}_i, \mathbf{s}_j) = d_\epsilon(\mathbf{s}_i, \mathbf{s}_j) \cdot v(\mathbf{s}_i, \mathbf{s}_j).$$

In what follows, we will analyze the continuum limit of the transition kernel k_ϵ for Gaussian diffusion kernel combined with different velocity kernels in Sections 4.2.1-4.2.3, and the analysis for kNN kernel in Section 4.2.4.

Remark 4.1. We remark that the diffusion kernel does not necessarily bring the diffusion in the final continuum limit of the RNA velocity models as we will see. The use of the name “diffusion” here only respects the convention that it will introduce diffusion type limit in the convergence analysis of graph Laplacians [52].

4.2.1 Continuum limit of cosine scheme

The transition probability matrix $P_\epsilon = (p_{ij})_{i,j=1:n_c}$ among cells through the Gaussian-cosine scheme is defined by

$$p_{ij} = \frac{k_\epsilon(\mathbf{s}_i, \mathbf{s}_j)}{\sum_{j=1}^{n_c} k_\epsilon(\mathbf{s}_i, \mathbf{s}_j)}, \tag{4.10}$$

which was utilized in [3].

To study the continuum limit of P_ϵ when the number of samples goes to infinity, we first study the operator \mathcal{G}_ϵ defined by

$$\mathcal{G}_\epsilon f(\mathbf{x}) = \frac{1}{\epsilon^{\frac{d}{2}}} \int_{\mathbb{R}^d} k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \tag{4.11}$$

We have the following lemma.

Lemma 4.1. *The operator \mathcal{G}_ϵ for Gaussian-cosine scheme has the expansion*

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \\ &= m_0 f(\mathbf{x}) + \sqrt{\epsilon} m_1 \mathcal{A} f(\mathbf{x}) + o(\sqrt{\epsilon}), \end{aligned}$$

where

$$\begin{aligned} m_0 &:= \mathcal{G}_\epsilon 1 = \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &= C_d \int_0^\infty r^{d-1} h(r^2) dr \int_{-\pi}^\pi |\sin \theta|^{d-2} g(\cos \theta) d\theta, \end{aligned} \tag{4.12}$$

$$m_1 := C_d \int_0^\infty r^d h(r^2) dr \int_{-\pi}^\pi \cos \theta |\sin \theta|^{d-2} g(\cos \theta) d\theta, \tag{4.13}$$

and

$$\mathcal{A}f(\mathbf{x}) = \|\nabla f(\mathbf{x})\| \cos\langle \mathbf{v}(\mathbf{x}), \nabla f(\mathbf{x}) \rangle = \hat{\mathbf{v}}(\mathbf{x}) \cdot \nabla f(\mathbf{x}), \quad \hat{\mathbf{v}}(\mathbf{x}) := \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

Here, $d > 1$, $C_d = S_d / \int_{-\pi}^{\pi} |\sin\theta|^{d-2} d\theta$ and S_d is the surface area of the d -dimensional unit sphere.

Proof. Let us assume $\mathbf{v}(\mathbf{x}) = \|\mathbf{v}\|(1, 0, \dots, 0)^T$ without loss of generality. The derived result can be transformed back to the original variables by substituting $(1, 0, \dots, 0)$ with the vector $\mathbf{v}(\mathbf{x}) / \|\mathbf{v}\|$.

For simplicity, we first consider the case $d = 2$. Define the polar coordinates

$$\begin{cases} y_1 = x_1 + r \cos\theta, \\ y_2 = x_2 + r \sin\theta, \end{cases}$$

where θ is the angle between $\mathbf{y} - \mathbf{x}$ and $\mathbf{v}(\mathbf{x})$. Then we obtain

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon} \int_0^\infty \int_{-\pi}^\pi rh \left(\frac{r^2}{\epsilon} \right) g(\cos\theta) f(r, \theta) d\theta dr \\ &\stackrel{r \rightarrow \sqrt{\epsilon}r}{=} \frac{1}{\epsilon} \int_0^\infty \int_{-\pi}^\pi \sqrt{\epsilon} rh(r^2) g(\cos\theta) f(\sqrt{\epsilon}r, \theta) d\theta d\sqrt{\epsilon}r \\ &= \int_0^\infty rh(r^2) \int_{-\pi}^\pi g(\cos\theta) f(\sqrt{\epsilon}r, \theta) d\theta dr \\ &= \int_{\epsilon^{\gamma-\frac{1}{2}}}^\infty + \int_0^{\epsilon^{\gamma-\frac{1}{2}}} \left(rh(r^2) \int_{-\pi}^\pi g(\cos\theta) f(\sqrt{\epsilon}r, \theta) d\theta \right) dr \\ &:= Q_1 + Q_2, \end{aligned} \tag{4.14}$$

where $0 < \gamma < \frac{1}{2}$. We have

$$\begin{aligned} Q_1 &= \int_{\epsilon^{\gamma-\frac{1}{2}}}^\infty rh(r^2) \int_{-\pi}^\pi g(\cos\theta) f(\sqrt{\epsilon}r, \theta) d\theta dr \\ &\leq C \exp(-\epsilon^{2\gamma-1}) = o(\sqrt{\epsilon}), \end{aligned}$$

where C depends on $\|f\|_\infty$ and $\|g\|_\infty$ and we utilized the exponential decay of $h(\cdot)$ and the inequality $\int_a^\infty xe^{-x^2} dx \leq \frac{1}{2}e^{-a^2}$ for $a > 0$.

For the integral in Q_2 , by Taylor expansion

$$f(\sqrt{\epsilon}r, \theta) = f(0, \theta) + \sqrt{\epsilon}r \frac{\partial f}{\partial r} \Big|_{(0, \theta)} + o(\sqrt{\epsilon}),$$

we get

$$\begin{aligned}
 Q_2 &= \int_0^{\epsilon^{\gamma-\frac{1}{2}}} rh(r^2) \int_{-\pi}^{\pi} g(\cos\theta) f(\sqrt{\epsilon}r, \theta) d\theta dr \\
 &= \int_0^{\epsilon^{\gamma-\frac{1}{2}}} rh(r^2) \int_{-\pi}^{\pi} g(\cos\theta) \left(f(0, \theta) + \sqrt{\epsilon}r \frac{\partial f}{\partial r} \Big|_{(0, \theta)} \right) d\theta dr + o(\sqrt{\epsilon}) \\
 &= \int_0^{\infty} rh(r^2) \int_{-\pi}^{\pi} g(\cos\theta) \left(f(0, \theta) + \sqrt{\epsilon}r \frac{\partial f}{\partial r} \Big|_{(0, \theta)} \right) d\theta dr + o(\sqrt{\epsilon}) \\
 &= m_0 f(x) + \sqrt{\epsilon} \int_0^{\infty} r^2 h(r^2) \int_{-\pi}^{\pi} g(\cos\theta) \frac{\partial f}{\partial r} \Big|_{(0, \theta)} d\theta dr + o(\sqrt{\epsilon}),
 \end{aligned}$$

where the extension from the integral domain $[0, \epsilon^{\gamma-\frac{1}{2}}]$ to $[0, \infty)$ will only introduce an exponentially small term by similar argument in estimating Q_1 .

For the $\mathcal{O}(\sqrt{\epsilon})$ term, note the relation

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x_1} \cos\theta + \frac{\partial f}{\partial x_2} \sin\theta$$

between polar and Euclidean coordinates, we get

$$\begin{aligned}
 &\int_{-\pi}^{\pi} g(\cos\theta) \frac{\partial f}{\partial r} \Big|_{(0, \theta)} d\theta \\
 &= \int_{-\pi}^{\pi} g(\cos\theta) \left(\frac{\partial f}{\partial x_1} \Big|_x \cos\theta + \frac{\partial f}{\partial x_2} \Big|_x \sin\theta \right) d\theta \\
 &= \|\nabla f(x)\| \cos\langle v(x), \nabla f(x) \rangle \int_{-\pi}^{\pi} g(\cos\theta) \cos\theta d\theta.
 \end{aligned} \tag{4.15}$$

This finishes the proof of the case $d=2$.

In high dimensions, the derivation is similar. We may consider the coordinate transformation from (y_1, y_2, \dots, y_d) to $(r, \theta, z_2, \dots, z_{d-2})$ defined by

$$\begin{cases}
 y_1 = x_1 + r \cos\theta, \\
 y_2 = x_2 + r \sin\theta \cdot z_2, \\
 \vdots \\
 y_{d-1} = x_{d-1} + r \sin\theta \cdot z_{d-1}, \\
 y_d = x_d + r \sin\theta \cdot \sqrt{1 - \sum_{i=2}^{d-1} z_i^2},
 \end{cases} \tag{4.16}$$

where $r \geq 0$, $-\pi \leq \theta \leq \pi$, $\sum_{i=2}^{d-1} z_i^2 \leq 1$. Denote

$$\mathbf{z} = (z_2, \dots, z_{d-1}), \quad z_d = \left(1 - \sum_{i=2}^{d-1} z_i^2 \right)^{\frac{1}{2}}.$$

Then the Jacobian

$$\frac{\partial(y_1, \dots, y_d)}{\partial(r, \theta, z_2, \dots, z_{d-1})} = \frac{(-1)^d (\sin \theta)^{d-2} r^{d-1}}{z_d}.$$

Therefore, we obtain

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{d/2}} \int_0^\infty \int_{-\pi}^\pi \int_{\|z\| \leq 1} h\left(\frac{r^2}{\epsilon}\right) g(\cos \theta) f(r, \theta, \mathbf{z}) \frac{|\sin \theta|^{d-2} r^{d-1}}{z_d} dz d\theta dr \\ &= \int_0^\infty r^{d-1} h(r^2) dr \int_{-\pi}^\pi |\sin \theta|^{d-2} g(\cos \theta) d\theta \int_{\|z\| \leq 1} \frac{1}{z_d} f(\sqrt{\epsilon} r, \theta, \mathbf{z}) dz. \end{aligned}$$

With Taylor expansion

$$f(\sqrt{\epsilon} r, \theta, \mathbf{z}) = f(0, \theta, \mathbf{z}) + \sqrt{\epsilon} r \frac{\partial f}{\partial r} \Big|_{(0, \theta, \mathbf{z})} + o(\sqrt{\epsilon})$$

and similar techniques in estimating (4.14) by Laplace asymptotics, we get

$$\mathcal{G}_\epsilon f(\mathbf{x}) = m_0 f(\mathbf{x}) + \sqrt{\epsilon} f_1(\mathbf{x}) + o(\sqrt{\epsilon}), \quad (4.17)$$

where

$$f_1(\mathbf{x}) = \int_0^\infty r^d h(r^2) dr \int_{-\pi}^\pi |\sin \theta|^{d-2} g(\cos \theta) d\theta \int_{\|z\| \leq 1} \frac{1}{z_d} \frac{\partial f}{\partial r} \Big|_{(0, \theta, \mathbf{z})} dz.$$

From the relation

$$\frac{\partial f}{\partial r} \Big|_{(0, \theta, \mathbf{z})} = \sum_{i=1}^d \frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial r} = \frac{\partial f}{\partial y_1} \Big|_x \cos \theta + \frac{\partial f}{\partial y_2} \Big|_x \sin \theta z_2 + \dots + \frac{\partial f}{\partial y_d} \Big|_x \sin \theta z_d$$

and noting the integral of odd functions with respect to θ vanishes, we can simplify the $\mathcal{O}(\sqrt{\epsilon})$ term as

$$\begin{aligned} f_1(\mathbf{x}) &= \int_0^\infty r^d h(r^2) dr \int_{-\pi}^\pi C_d |\sin \theta|^{d-2} \cos \theta g(\cos \theta) d\theta \cdot \frac{\partial f}{\partial y_1} \Big|_x \\ &= m_1 \mathcal{A} f(\mathbf{x}), \end{aligned}$$

where we used the result $\int_{\|z\| \leq 1} z_d^{-1} dz = C_d$. \square

Given the sample probability density $q(\mathbf{x})$, then the weighted graph Laplacian has the form

$$p_\epsilon(\mathbf{x}, \mathbf{y}) = \frac{k_\epsilon(\mathbf{x}, \mathbf{y})}{d_\epsilon(\mathbf{x})}, \quad d_\epsilon(\mathbf{x}) = \int k_\epsilon(\mathbf{x}, \mathbf{y}) q(\mathbf{y}) d\mathbf{y}. \quad (4.18)$$

Define the operator

$$\mathcal{P}_\epsilon f(\mathbf{x}) = \int p_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) q(\mathbf{y}) d\mathbf{y}$$

and the generator

$$\mathcal{L}_\epsilon = \frac{\mathcal{P}_\epsilon - I}{\sqrt{\epsilon}}.$$

Theorem 4.1 (Continuum Limit of the Gaussian-Cosine Scheme). *For the Gaussian-Cosine scheme, we have the limit infinitesimal generator*

$$\lim_{\epsilon \rightarrow 0^+} \mathcal{L}_\epsilon f = \mathcal{L}f := \frac{m_1}{m_0} \hat{v}(\mathbf{x}) \cdot \nabla f(\mathbf{x}), \quad \hat{v}(\mathbf{x}) := \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

Proof. According to Lemma 4.1, we have

$$\begin{aligned} \mathcal{P}_\epsilon f(\mathbf{x}) &= \int p_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) q(\mathbf{y}) d\mathbf{y} \\ &= \frac{\int k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) q(\mathbf{y}) d\mathbf{y}}{\int k_\epsilon(\mathbf{x}, \mathbf{y}) q(\mathbf{y}) d\mathbf{y}} = \frac{\mathcal{G}_\epsilon(fq)(\mathbf{x})}{\mathcal{G}_\epsilon q(\mathbf{x})} \\ &= \frac{m_0 f(\mathbf{x}) q(\mathbf{x}) + \sqrt{\epsilon} m_1 \hat{v}(\mathbf{x}) \cdot \nabla(fq)(\mathbf{x}) + o(\sqrt{\epsilon})}{m_0 q(\mathbf{x}) + \sqrt{\epsilon} m_1 \hat{v}(\mathbf{x}) \cdot \nabla q(\mathbf{x}) + o(\sqrt{\epsilon})} \\ &= f(\mathbf{x}) + \sqrt{\epsilon} \frac{m_1}{m_0} \hat{v}(\mathbf{x}) \cdot \nabla f(\mathbf{x}) + o(\sqrt{\epsilon}). \end{aligned}$$

Then the theorem follows obviously. \square

Remark 4.2. Theorem 4.1 implies that in the continuum limit, if $m_1 \neq 0$ (it holds for general $g(\cdot)$), then the Markov semigroup defined by P_ϵ corresponds to the ODE dynamics

$$\frac{d\mathbf{x}}{dt} = \hat{v}(\mathbf{x}). \quad (4.19)$$

In this setup, the source and sink states defined by tracing the integral curves of (4.19) will be the same as those obtained by solving

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{x}), \quad (4.20)$$

since the change of the magnitude of the velocity in (4.20) does not affect the shape of the integral curves but a reparameterization. However, the transition rule defined by (4.10) is not effective on analyzing the landscape and transition behavior among the pluripotent and differentiated cells, which will be addressed in the inner-product scheme.

Remark 4.3. The implementation in [3] actually centers both δ_{ij} and v_i . So the cosine kernel considered in their paper is equivalent to the correlation kernel below but not the cosine kernel considered here.

4.2.2 Continuum limit of correlation scheme

The correlation scheme has been utilized in [29, 41]. We have similar expansion to the operator \mathcal{G}_ϵ for Gaussian-Correlation scheme.

Lemma 4.2. *The operator \mathcal{G}_ϵ for Gaussian-Correlation scheme has the expansion*

$$\begin{aligned}\mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \\ &= m_0 f(\mathbf{x}) + \sqrt{\epsilon} m_1 \mathcal{A}f(\mathbf{x}) + o(\sqrt{\epsilon}),\end{aligned}$$

where

$$\begin{aligned}m_0 &:= \mathcal{G}_\epsilon 1 = \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &= \begin{cases} C_{d-1} \int_{\mathbb{R}} \int_0^\infty r^{d-2} h(r^2 + \mu^2) dr d\mu \int_{-\pi}^\pi |\sin\theta|^{d-3} g(\cos\theta) d\theta, & d > 2, \\ \int_{\mathbb{R}} \int_0^\infty h(r^2 + \mu^2) [g(1) + g(-1)] dr d\mu, & d = 2, \end{cases} \\ m_1 &:= \begin{cases} C_{d-1} \int_{\mathbb{R}} \int_0^\infty r^{d-1} h(r^2 + \mu^2) dr d\mu \int_{-\pi}^\pi \cos\theta |\sin\theta|^{d-3} g(\cos\theta) d\theta, & d > 2, \\ \int_{\mathbb{R}} \int_0^\infty r h(r^2 + \mu^2) [g(1) - g(-1)] dr d\mu, & d = 2, \end{cases}\end{aligned}$$

and

$$\mathcal{A}f(\mathbf{x}) = \widehat{\mathcal{P}_1 \mathbf{v}}(\mathbf{x}) \cdot \nabla f(\mathbf{x}), \quad \widehat{\mathcal{P}_1 \mathbf{v}} := \frac{\mathcal{P}_1 \mathbf{v}}{\|\mathcal{P}_1 \mathbf{v}\|}.$$

Proof. Define projection operator

$$\mathcal{P}_n \mathbf{x} := (I - \hat{\mathbf{n}} \otimes \hat{\mathbf{n}}) \cdot \mathbf{x},$$

where \mathbf{n} is the normal vector and $\hat{\mathbf{n}} = \mathbf{n} / \|\mathbf{n}\|$. Then the correlation between vectors \mathbf{x} and \mathbf{y} is

$$\text{corr}(\mathbf{x}, \mathbf{y}) := \cos\langle \mathcal{P}_1 \mathbf{x}, \mathcal{P}_1 \mathbf{y} \rangle,$$

where $\mathbf{1} = (1, \dots, 1)^\top$. It is not difficult to find that the correlation between \mathbf{x} and \mathbf{y} is invariant with respect to the rotations in the hyperplane which is normal to $\mathbf{1}$. With this observation, we utilize a new coordinate system which maps

$$\frac{1}{\sqrt{d}} \mathbf{1} \mapsto \mathbf{1}_d = (0, \dots, 0, 1)^\top, \quad \frac{\mathcal{P}_1 \mathbf{v}}{\|\mathcal{P}_1 \mathbf{v}\|} \mapsto (1, 0, \dots, 0)^\top.$$

The derived result can be transformed back to the original variable like Lemma 4.1 in a straightforward way.

In this new coordinate system, the analysis of correlation scheme is similar to the cosine scheme for its first $d-1$ components. Then we may consider the coordinate transformation below from (y_1, y_2, \dots, y_d) to $(r, \mu, \theta, z_2, \dots, z_{d-2})$ when $d > 2$ (the case $d = 2$ is

easier and not necessary to do the coordinate transformation since the correlation will be ± 1 in this case):

$$\begin{cases} y_1 = x_1 + r \cos \theta, \\ y_2 = x_2 + r \sin \theta \cdot z_2, \\ \vdots \\ y_{d-2} = x_{d-2} + r \sin \theta \cdot z_{d-2}, \\ y_{d-1} = x_{d-1} + r \sin \theta \cdot \sqrt{1 - \sum_{i=2}^{d-2} z_i^2}, \\ y_d = x_d + \mu, \end{cases}$$

where $r \geq 0, \mu \in \mathbb{R}, -\pi \leq \theta \leq \pi, \sum_{i=2}^{d-2} z_i^2 \leq 1$. Denote

$$\mathbf{z} = (z_2, \dots, z_{d-2}), \quad z_{d-1} = \left(1 - \sum_{i=2}^{d-2} z_i^2\right)^{\frac{1}{2}}.$$

Then the Jacobian

$$\frac{\partial(y_1, \dots, y_d)}{\partial(r, \mu, \theta, z_2, \dots, z_{d-2})} = \frac{(-1)^{d-1} (\sin \theta)^{d-3} r^{d-2}}{z_{d-1}}.$$

Therefore, we obtain

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{d/2}} \int_{\mathbb{R}} \int_0^\infty \int_{-\pi}^\pi \int_{\|\mathbf{z}\| \leq 1} h\left(\frac{r^2 + \mu^2}{\epsilon}\right) g(\cos \theta) f(r, \mu, \theta, \mathbf{z}) \\ &\quad \cdot \frac{|\sin \theta|^{d-3} r^{d-2}}{z_{d-1}} dz d\theta dr d\mu \\ &= \int_0^\infty r^{d-2} h(r^2 + \mu^2) dr d\mu \int_{-\pi}^\pi |\sin \theta|^{d-3} g(\cos \theta) d\theta \\ &\quad \cdot \int_{\|\mathbf{z}\| \leq 1} \frac{1}{z_{d-1}} f(\sqrt{\epsilon} r, \sqrt{\epsilon} \mu, \theta, \mathbf{z}) dz. \end{aligned}$$

With Taylor expansion

$$f(\sqrt{\epsilon} r, \sqrt{\epsilon} \mu, \theta, \mathbf{z}) = f(0, 0, \theta, \mathbf{z}) + \sqrt{\epsilon} r \frac{\partial f}{\partial r} \Big|_{(0,0,\theta,\mathbf{z})} + \sqrt{\epsilon} \mu \frac{\partial f}{\partial \mu} \Big|_{(0,0,\theta,\mathbf{z})} + o(\sqrt{\epsilon})$$

and similar techniques in estimating (4.14) by Laplace asymptotics, we get

$$\mathcal{G}_\epsilon f(\mathbf{x}) = m_0 f(\mathbf{x}) + \sqrt{\epsilon} f_1(\mathbf{x}) + o(\sqrt{\epsilon}), \tag{4.21}$$

where

$$\begin{aligned} f_1(\mathbf{x}) &= \int_{\mathbb{R}} \int_0^\infty r^{d-1} h(r^2 + \mu^2) dr d\mu \int_{-\pi}^\pi |\sin \theta|^{d-3} g(\cos \theta) d\theta \\ &\quad \cdot \int_{\|\mathbf{z}\| \leq 1} \frac{1}{z_{d-1}} \frac{\partial f}{\partial r} \Big|_{(0,0,\theta,\mathbf{z})} dz. \end{aligned}$$

Here we have ignored the odd function with respect to μ , then according to the chain rule

$$\left. \frac{\partial f}{\partial r} \right|_{(0,0,\theta,z)} = \sum_{i=1}^d \frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial r},$$

and noting the integral of odd functions with respect to θ vanishes, we can simplify the $\mathcal{O}(\sqrt{\epsilon})$ term as

$$\begin{aligned} f_1(\mathbf{x}) &= \int_{\mathbb{R}} \int_0^{\infty} r^{d-1} h(r^2 + \mu^2) dr d\mu \int_{-\pi}^{\pi} C_{d-1} \cos\theta |\sin\theta|^{d-3} g(\cos\theta) d\theta \cdot \left. \frac{\partial f}{\partial y_1} \right|_{\mathbf{x}} \\ &= m_1 \mathcal{A}f(\mathbf{x}). \end{aligned}$$

The proof is done. □

Then, similar to Theorem 4.1, we get the generator \mathcal{L} .

Theorem 4.2 (Continuum Limit of the Gaussian-Correlation Scheme). *For the Gaussian-correlation scheme, we have the limit infinitesimal generator*

$$\lim_{\epsilon \rightarrow 0^+} \mathcal{L}_{\epsilon} f = \mathcal{L}f := \frac{m_1}{m_0} \widehat{\mathcal{P}}_1 \mathbf{v}(\mathbf{x}) \cdot \nabla f(\mathbf{x}).$$

Remark 4.4. Theorem 4.2 implies that in the continuum limit of correlation scheme, if $m_1 \neq 0$, the Markov semigroup defined by P_{ϵ} corresponds to the ODE dynamics

$$\frac{d\mathbf{x}}{dt} = \widehat{\mathcal{P}}_1 \mathbf{v}(\mathbf{x}). \quad (4.22)$$

This means the effective velocity of the correlation scheme is

$$\tilde{\mathbf{v}}(\mathbf{x}) = (I - \mathbf{1} \cdot \mathbf{1}^T) \cdot \mathbf{v}(\mathbf{x})$$

which will introduce bias into the final result in the identification of root and ending cells. We speculate that the correlation scheme has the possibility to give undesired result in the sense that it is not exactly respect to the inferred velocity.

4.2.3 Continuum limit of inner-product scheme

The inner-product scheme is constructed similar to diffusion map [9]. Related idea and methodology has been utilized to analyze the scRNA-seq data analysis [46, 57, 68].

Given the sample probability density $q(\mathbf{x})$, we define a new kernel

$$k_{\epsilon}^{(\alpha)}(\mathbf{x}, \mathbf{y}) = \frac{k_{\epsilon}(\mathbf{x}, \mathbf{y})}{q_{\epsilon}^{\alpha}(\mathbf{x}) q_{\epsilon}^{\alpha}(\mathbf{y})}, \quad q_{\epsilon}(\mathbf{x}) = \int k_{\epsilon}(\mathbf{x}, \mathbf{y}) q(\mathbf{y}) d\mathbf{y}. \quad (4.23)$$

Then we apply the weighted graph Laplacian normalization to this kernel by

$$p_{\epsilon, \alpha}(\mathbf{x}, \mathbf{y}) = \frac{k_{\epsilon}^{(\alpha)}(\mathbf{x}, \mathbf{y})}{d_{\epsilon}^{(\alpha)}(\mathbf{x})}, \quad d_{\epsilon}^{(\alpha)}(\mathbf{x}) = \int k_{\epsilon}^{(\alpha)}(\mathbf{x}, \mathbf{y}) q(\mathbf{y}) d\mathbf{y}. \quad (4.24)$$

The application of the above construction to the RNA velocity data is straightforward by replacing the density $q(\mathbf{x})$ with the empirical data distribution of the cell states, i.e.

$$q(\mathbf{x}) = \frac{1}{n_c} \sum_{j=1}^{n_c} \delta(\mathbf{x} - \mathbf{s}_j).$$

To study the continuum limit in space and time, we define the operator

$$\mathcal{P}_{\epsilon, \alpha} f(\mathbf{x}) = \int p_{\epsilon, \alpha}(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) q(\mathbf{y}) d\mathbf{y}$$

and the generator

$$\mathcal{L}_{\epsilon, \alpha} = \frac{\mathcal{P}_{\epsilon, \alpha} - I}{\epsilon}.$$

First let us define \mathcal{G}_ϵ as in (4.11). We have the following lemma.

Lemma 4.3. *The operator \mathcal{G}_ϵ for Gaussian-Inner-product scheme has the expansion*

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \\ &= m_0 g(0) f(\mathbf{x}) + \epsilon \frac{m_2}{2} \mathcal{A}f(\mathbf{x}) + o(\epsilon), \end{aligned}$$

where the moments

$$m_0 := \int h(\|\mathbf{y}\|^2) d\mathbf{y}, \quad m_2 := \frac{1}{d} \int \|\mathbf{y}\|^2 h(\|\mathbf{y}\|^2) d\mathbf{y}$$

and

$$\mathcal{A}f(\mathbf{x}) = \|\mathbf{v}(\mathbf{x})\|^2 g''(0) f(\mathbf{x}) + 2g'(0) \mathbf{v}(\mathbf{x})^T \nabla f(\mathbf{x}) + g(0) \Delta f(\mathbf{x}).$$

Proof. The key idea is to utilize the Laplace asymptotics similar as in estimating (4.14). With inner-product scheme, we have

$$k_\epsilon(\mathbf{x}, \mathbf{y}) = h\left(\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\epsilon}\right) g((\mathbf{y} - \mathbf{x})^T \mathbf{v}(\mathbf{x})).$$

Make the decomposition

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{\frac{d}{2}}} \int_{\|\mathbf{x} - \mathbf{y}\| > \epsilon^\gamma} k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \\ &\quad + \frac{1}{\epsilon^{\frac{d}{2}}} \int_{\|\mathbf{x} - \mathbf{y}\| \leq \epsilon^\gamma} k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \\ &:= Q_1 + Q_2, \end{aligned}$$

where $0 < \gamma < \frac{1}{2}$. Let $\mathbf{y} = \mathbf{x} + \sqrt{\epsilon} \mathbf{z}$. The term

$$\begin{aligned} Q_1 &= \int_{\|\mathbf{z}\| > \epsilon^{\gamma - \frac{1}{2}}} h(\|\mathbf{z}\|^2) g(\sqrt{\epsilon} \mathbf{v}(\mathbf{x})^T \mathbf{z}) f(\mathbf{x} + \sqrt{\epsilon} \mathbf{z}) d\mathbf{z} \\ &\leq C \epsilon^{\frac{1}{2} - \gamma} \exp(-\epsilon^{2\gamma - 1}) = o(\epsilon) \end{aligned}$$

where C depends on $\|f\|_\infty$ and $\|g\|_\infty$, and we utilized the exponential decay of $h(\cdot)$ and the inequality $\int_a^\infty e^{-x^2} dx \leq (2a)^{-1} e^{-a^2}$ for $a > 0$.

For the integrand in Q_2 , we have

$$g(\sqrt{\epsilon}v^T z)f(x + \sqrt{\epsilon}z) = g(0)f(x) + \sqrt{\epsilon}P_1 + \epsilon P_2 + o(\epsilon)$$

by Taylor expansion, where

$$\begin{aligned} P_1 &= z^T (g'(0)f(x)v + g(0)\nabla f(x)), \\ P_2 &= \frac{1}{2}\text{Tr}\left(zz^T \cdot \left(2g'(0)\nabla f(x)v^T + g''(0)f(x)vv^T + g(0)\nabla^2 f(x)\right)\right). \end{aligned}$$

Here $\text{Tr}(\cdot)$ is the matrix trace operation.

By the rotation symmetry of $\|z\|^2$ and the fact that the integral of an odd function is zero, we obtain

$$\begin{aligned} Q_2 &= \int_{\|z\| \leq \epsilon^{\gamma-\frac{1}{2}}} \left(h(\|z\|^2)g(0)f(x) + \frac{1}{2d}\epsilon h(\|z\|^2)\|z\|^2 \mathcal{A}f(x) \right) dz + o(\epsilon) \\ &= \int \left(h(\|z\|^2)g(0)f(x) + \frac{1}{2d}\epsilon h(\|z\|^2)\|z\|^2 \mathcal{A}f(x) \right) dz + o(\epsilon) \\ &= m_0 g(0)f(x) + \epsilon \frac{m_2}{2} \mathcal{A}f(x) + o(\epsilon), \end{aligned}$$

where the extension of the integral domain from $\|z\| \leq \epsilon^{\gamma-\frac{1}{2}}$ to the whole space will only introduce an $o(\epsilon)$ term by similar argument in estimating Q_1 . \square

Theorem 4.3 (Continuum Limit of the Gaussian-Inner-product Scheme). *For the Gaussian-inner-product scheme, we have the limit infinitesimal generator*

$$\lim_{\epsilon \rightarrow 0^+} \mathcal{L}_{\epsilon, \alpha} f = \mathcal{L}_\alpha f := \frac{m_2}{m_0} \left(\frac{1}{2} \Delta f + \left((1-\alpha) \frac{\nabla q}{q} + \frac{g'(0)}{g(0)} v \right) \cdot \nabla f \right). \quad (4.25)$$

Proof. First note that $p_{\epsilon, \alpha}$ in (4.24) is invariant under any multiplicative scaling applied to $k_\epsilon^{(\alpha)}$. So we can assume in what follows that the normalization $\epsilon^{d/2}$ is implicitly contained in $k_\epsilon(x, y)$, which will not affect the final result.

According to Lemma 4.3, we have the asymptotics

$$q_\epsilon(x) = m_0 g(0)q(x) + \epsilon \frac{m_2}{2} \mathcal{A}q(x) + o(\epsilon)$$

for q_ϵ defined in (4.23). Correspondingly,

$$q_\epsilon^{-\alpha}(x) = (m_0 g(0)q(x))^{-\alpha} (1 + \epsilon \mathcal{B}q(x) + o(\epsilon)), \quad \mathcal{B}q(x) = -\frac{\alpha m_2}{2m_0 g(0)q(x)} \mathcal{A}q(x).$$

Define

$$\mathcal{G}_\epsilon^{(\alpha)} \phi(x) = \int k_\epsilon^{(\alpha)}(x, y) \phi(y) q(y) dy.$$

Then

$$\begin{aligned}
 \mathcal{G}_\epsilon^{(\alpha)}\phi(\mathbf{x}) &= q_\epsilon^{-\alpha} \int k_\epsilon(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) q_\epsilon^{-\alpha}(\mathbf{y}) q(\mathbf{y}) d\mathbf{y} \\
 &= q_\epsilon^{-\alpha} \int k_\epsilon(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) \frac{q^{1-\alpha}(\mathbf{y})}{(m_0 g(0))^\alpha} (1 + \epsilon \mathcal{B}q(\mathbf{y}) + o(\epsilon)) d\mathbf{y} \\
 &= q_\epsilon^{-\alpha}(\mathbf{x}) \phi(\mathbf{x}) (m_0 g(0) q(\mathbf{x}))^{1-\alpha} (1 + \epsilon \mathcal{B}q(\mathbf{x}) + o(\epsilon)) \\
 &\quad + \epsilon q_\epsilon^{-\alpha}(\mathbf{x}) \frac{m_2}{2} (m_0 g(0))^{-\alpha} \mathcal{A}(\phi(\mathbf{x}) q^{1-\alpha}(\mathbf{x})) + o(\epsilon).
 \end{aligned} \tag{4.26}$$

So we obtain

$$\begin{aligned}
 \mathcal{P}_{\epsilon, \alpha} f(\mathbf{x}) &= \int p_{\epsilon, \alpha}(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) q(\mathbf{y}) d\mathbf{y} = \frac{\mathcal{G}_\epsilon^{(\alpha)} f(\mathbf{x})}{\mathcal{G}_\epsilon^{(\alpha)} 1(\mathbf{x})} \\
 &= f(\mathbf{x}) + \frac{\epsilon m_2}{2 m_0 g(0) q^{1-\alpha}(\mathbf{x})} \left(\mathcal{A}(f(\mathbf{x}) q^{1-\alpha}(\mathbf{x})) - f(\mathbf{x}) \mathcal{A} q^{1-\alpha}(\mathbf{x}) \right) + o(\epsilon) \\
 &= f(\mathbf{x}) + \epsilon \mathcal{L}_\alpha f(\mathbf{x}) + o(\epsilon).
 \end{aligned} \tag{4.27}$$

Finally we get

$$\lim_{\epsilon \rightarrow 0^+} \mathcal{L}_{\epsilon, \alpha} f(\mathbf{x}) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\mathcal{P}_{\epsilon, \alpha} f(\mathbf{x}) - f(\mathbf{x})) = \mathcal{L}_\alpha f(\mathbf{x}).$$

The proof is done. □

Corollary 4.1. *If we choose $h(\cdot), g(\cdot)$ such that $m_2 = 2, m_0 = 1$ and $g'(0) = \frac{1}{2}g(0)$, then*

$$\mathcal{L}_\alpha f(\mathbf{x}) = \left(\mathbf{v}(\mathbf{x}) + 2(1-\alpha) \frac{\nabla q}{q} \right) \cdot \nabla f(\mathbf{x}) + \Delta f(\mathbf{x}),$$

which is the generator of the stochastic differential equations (SDEs)

$$d\mathbf{x}(t) = \left(\mathbf{v}(\mathbf{x}) + 2(1-\alpha) \frac{\nabla q}{q}(\mathbf{x}) \right) dt + \sqrt{2} d\mathbf{w}(t), \tag{4.28}$$

where $\mathbf{w}(t)$ is the standard Brownian motion with mean 0 and covariance function $\mathbb{E}(\mathbf{w}(t)\mathbf{w}(s)) = \min(t, s)$. In general case, similar SDE holds with suitable constants $m_0, m_2, g(0)$ and $g'(0)$.

Define the data potential

$$V(\mathbf{x}) = -\log q(\mathbf{x}).$$

Then the SDEs (4.28) becomes

$$d\mathbf{x}(t) = (\mathbf{v}(\mathbf{x}) - 2(1-\alpha) \nabla V(\mathbf{x})) dt + \sqrt{2} d\mathbf{w}(t),$$

where the drift term is composed of the gradient part $-\nabla V(\mathbf{x})$ from the scRNA-seq sampling distribution, and the non-gradient part from the RNA velocity $\mathbf{v}(\mathbf{x})$. Specifically, if we choose $\alpha = 1$, then the transition probability will not depend on the data potential $V(\mathbf{x})$ in the infinite samples limit. This structure opens the way of studying the non-equilibrium steady state and landscape theory for cell developments with scRNA-seq experimental data [18, 46, 57, 65, 67, 68].

Remark 4.5. In [4], the authors proposed a class of local kernels to approximate the SDE (4.2), which is consistent with our analysis of the inner-product scheme. Note that one typical choice of local kernel is $\exp(-\frac{\|x-y-\epsilon v(x)\|^2}{2\epsilon}) = \exp(-\frac{\|x-y\|^2}{2\epsilon}) \exp(-(\mathbf{y}-x)^T v(x)) \exp(-\frac{1}{2}\epsilon \|v(x)\|^2)$, whose difference with $k_\epsilon(x, y)$ in inner-product scheme is up to $\mathcal{O}(\epsilon)$ in $g(\cdot)$, incurring no difference in the limit of infinitesimal generator.

Remark 4.6. In [29], the transition probabilities among cells were constructed through averaging a RNA velocity-based dynamics and a diffusion-based dynamics. For instance, we consider the transition matrix $\mathcal{P}_{\epsilon, \alpha}^\omega = \omega \mathcal{P}_{\epsilon, \alpha}^i + (1-\omega) \mathcal{P}_{\epsilon, \alpha}^d$, where $\mathcal{P}_{\epsilon, \alpha}^i$ is the transition probability matrix from inner-product kernel, $\mathcal{P}_{\epsilon, \alpha}^d$ from pure diffusion kernel and $0 < \omega < 1$ is the weight of averaging. Next we provide an understanding of $\mathcal{P}_{\epsilon, \alpha}^\omega$ based on the decomposition of SDE (4.2) into the equilibrium and non-equilibrium parts [1, 56, 61, 67].

For simplicity, let us assume that the ground-truth dynamics underlying cell-fate development is $d\mathbf{x}(t) = v(\mathbf{x})dt + \sqrt{2}d\mathbf{w}(t)$ and the distribution of all data points approximates its steady-state distribution $p^{ss}(\mathbf{x})$, which can be guaranteed by the ergodicity of dynamics. With the defined potential landscape $\phi = -\ln p^{ss}$, the velocity term can be decomposed [56] as $v(\mathbf{x}) = -\nabla\phi(\mathbf{x}) + \ell(\mathbf{x})$, $\ell(\mathbf{x}) = J^{ss}(\mathbf{x})/p^{ss}(\mathbf{x})$, where $J^{ss} = v p^{ss} - \nabla p^{ss}$ is the steady-state probability flux satisfying $\nabla \cdot J^{ss} = 0$. In terms of the statistical physics interpretation [18], the gradient term $-\nabla\phi$ can be viewed as the equilibrium part of velocity, and the curl-like term $\ell(\mathbf{x})$ is the non-equilibrium part.

With the assumptions above, we apply our results in Corollary 4.1 to the infinitesimal generator of averaging dynamics defined by $\mathcal{P}_{\epsilon, \alpha}^\omega$, and conclude that its continuum limit has the form

$$d\mathbf{x}(t) = [-(\omega + 2(1-\alpha))\nabla\phi(\mathbf{x}) + \omega\ell(\mathbf{x})]dt + \sqrt{2}d\mathbf{w}(t),$$

and the relative proportion of non-equilibrium part $\omega/(\omega + 2(1-\alpha))$ is an increasing function of ω if $0 < \alpha < 1$. Therefore the introduction of weight ω can be understood as tuning the relative weight of equilibrium and non-equilibrium parts of the RNA velocity.

4.2.4 Implications of kNN kernels

Sometimes people prefer kNN diffusion kernel with less computational effort due to the sparsity of transition matrix. For example, [3] used kNN-cosine kernel and [41] used kNN-correlation kernel. We will show that for kNN diffusion kernel, the conclusions above still hold with slight difference.

Lemma 4.4. *Suppose $q(\mathbf{x}) > 0$ everywhere. Let $r(\mathbf{x}) = (q(\mathbf{x}))^{-1/d}$ be the location-dependent bandwidth function. The operator \mathcal{G}_ϵ for kNN-cosine scheme has the expansion*

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \\ &= m_0(\mathbf{x}) f(\mathbf{x}) + \sqrt{\epsilon} m_1(\mathbf{x}) \mathcal{A} f(\mathbf{x}) + o(\sqrt{\epsilon}), \end{aligned}$$

where

$$\begin{aligned} m_0(\mathbf{x}) &:= \mathcal{G}_\epsilon 1 = \frac{1}{\epsilon^{\frac{d}{2}}} \int k_\epsilon(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &= C_d \int_0^{r(\mathbf{x})} r^{d-1} dr \int_{-\pi}^\pi |\sin\theta|^{d-2} g(\cos\theta) d\theta, \end{aligned} \tag{4.29}$$

$$m_1(\mathbf{x}) := C_d \int_0^{r(\mathbf{x})} r^d dr \int_{-\pi}^\pi \cos\theta |\sin\theta|^{d-2} g(\cos\theta) d\theta, \tag{4.30}$$

and

$$\mathcal{A}f(\mathbf{x}) = \|\nabla f(\mathbf{x})\| \cos\langle \mathbf{v}(\mathbf{x}), \nabla f(\mathbf{x}) \rangle = \hat{\mathbf{v}}(\mathbf{x}) \cdot \nabla f(\mathbf{x}), \quad \hat{\mathbf{v}}(\mathbf{x}) := \frac{\mathbf{v}}{\|\mathbf{v}\|},$$

i.e. we just replace constants m_0, m_1 in Lemma 4.1 with functions $m_0(\mathbf{x}), m_1(\mathbf{x})$.

Proof. The proof is similar to Lemma 4.1. We only use the case $d=2$ as an example. Now we have

$$\begin{aligned} \mathcal{G}_\epsilon f(\mathbf{x}) &= \int_0^{r_\epsilon(\mathbf{x})} r \int_{-\pi}^\pi g(\cos\theta) f(\sqrt{\epsilon}r, \theta) d\theta dr \\ &= \int_0^{r_\epsilon(\mathbf{x})} r \int_{-\pi}^\pi g(\cos\theta) \left(f(0, \theta) + \sqrt{\epsilon}r \frac{\partial f}{\partial r} \Big|_{(0, \theta)} \right) d\theta dr + o(\sqrt{\epsilon}) \\ &= \int_0^{r(\mathbf{x})} r \int_{-\pi}^\pi g(\cos\theta) \left(f(0, \theta) + \sqrt{\epsilon}r \frac{\partial f}{\partial r} \Big|_{(0, \theta)} \right) d\theta dr + o(\sqrt{\epsilon}) \\ &= m_0(\mathbf{x})f(\mathbf{x}) + \sqrt{\epsilon}m_1(\mathbf{x})\mathcal{A}f(\mathbf{x}) + o(\sqrt{\epsilon}), \end{aligned}$$

where we utilized the result $r_\epsilon(\mathbf{x}) \xrightarrow{a.s.} r(\mathbf{x}) + o(\sqrt{\epsilon})$ based on the Theorem A.2 in the Appendix. □

Now we can derive the limit infinitesimal generator.

Theorem 4.4 (Continuum Limit of the kNN-Cosine Scheme). *For the kNN-cosine scheme, we have the limit infinitesimal generator*

$$\lim_{\epsilon \rightarrow 0^+} \mathcal{L}_\epsilon f = \mathcal{L}f := \frac{m_1(\mathbf{x})}{m_0(\mathbf{x})} \hat{\mathbf{v}}(\mathbf{x}) \cdot \nabla f(\mathbf{x}) = c \cdot r(\mathbf{x}) \hat{\mathbf{v}}(\mathbf{x}) \cdot \nabla f(\mathbf{x}),$$

where c is a constant only related to $g(\cdot)$ and dimension d .

The results for kNN-correlation scheme is similar.

Remark 4.7. Based upon similar argument in Remark 4.2, the kNN-cosine scheme will give the same root and ending cells produced by Eq. (4.20) in the large n limit. The case with $q(\mathbf{x}) = 0$ is beyond our analysis framework. The kNN-inner-product scheme needs more delicate order analysis on the convergence of $r_\epsilon(\mathbf{x})$ to $r(\mathbf{x})$, which is considered in the Appendix.

Remark 4.8. In [29], the authors combined Gaussian diffusion kernel with kNN kernel, i.e.

$$d_\epsilon(\mathbf{s}_i, \mathbf{s}_j) = I \left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\epsilon r_\epsilon^2(\mathbf{s}_i)} \right) h \left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\epsilon} \right).$$

The overall analysis is similar. We only need to replace $m_0(\mathbf{x})$ in (4.29) with

$$m_0(\mathbf{x}) = C_d \int_0^{r(\mathbf{x})} r^{d-1} h(r^2) dr \int_{-\pi}^{\pi} |\sin\theta|^{d-2} g(\cos\theta) d\theta$$

by inserting $h(r^2)$ in the integral of r . The change of $m_1(\mathbf{x})$ is similar. The final limit infinitesimal generator is still

$$\mathcal{L}f(\mathbf{x}) = \frac{m_1(\mathbf{x})}{m_0(\mathbf{x})} \hat{\mathbf{v}}(\mathbf{x}) \cdot \nabla f(\mathbf{x}).$$

4.3 Finding root and ending cells

The root-and-ending cells finding algorithm has been proposed in [3, 29]. The aim of this section is to study its rationale through the continuum limit perspective.

From the derived continuum limits in Theorems 4.1 and 4.3, it is natural to identify the ending cells, i.e. the final differentiated cells, by selecting states with nonzero probability (or higher than a threshold in practice), from the invariant distribution π of the transition probability matrix $P_\epsilon = (p_{ij}^\epsilon)_{i,j=1:n_c}$:

$$\pi^\top = \pi^\top P_\epsilon.$$

In the limit case as studied in Theorem 4.1, these ending cells corresponding to the absorbing states of the ODE flow map

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{x}). \quad (4.31)$$

In this case, the limit ODE flow map is not irreducible, thus the invariant distributions are not unique in general. We should start from different, or random initial distributions at the beginning.

The identification of root cells, usually the stem cells, is more subtle. One first defines the backward transition probability matrix $\tilde{P}_\epsilon = (\tilde{p}_{ij}^\epsilon)_{i,j=1:n_c}$ by

$$\tilde{p}_{ij}^\epsilon = \frac{p_{ji}^\epsilon}{\sum_k p_{ki}^\epsilon}, \quad (4.32)$$

which is the row-normalization of P_ϵ^\top ; then identify the root cells by selecting states with probability above a threshold in the invariant distribution $\tilde{\pi}$ of \tilde{P}_ϵ .

To motivate the intuition of the above proposal, let us first study the continuous time limit of \tilde{P}_ϵ in the discrete states setup.

Theorem 4.5. *If A is the generator of a finite state Markov chain with transition probability matrix P_ϵ . Then, the generator of the backward transition probability matrix \tilde{P}_ϵ is*

$$\tilde{A} = A^T - \text{diag}(A^T \cdot \mathbf{1}). \tag{4.33}$$

Proof. By definition, we have

$$A_{ij} = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (p_{ij}^\epsilon - \delta_{ij}), \quad \text{i.e.} \quad p_{ij}^\epsilon = \delta_{ij} + \epsilon A_{ij} + o(\epsilon),$$

where δ_{ij} is the Kronecker's delta function. So we get

$$\begin{aligned} \tilde{A}_{ij} &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\tilde{p}_{ij}^\epsilon - \delta_{ij}) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \left(\frac{p_{ji}^\epsilon}{\sum_k p_{ki}^\epsilon} - \delta_{ij} \right) \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \left(\frac{\epsilon A_{ji} + \delta_{ji}}{\sum_k (\epsilon A_{ki} + \delta_{ki})} - \delta_{ij} \right) \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \frac{\epsilon A_{ji} + \delta_{ji} - \delta_{ij} \sum_k (\epsilon A_{ki} + \delta_{ki})}{\sum_k (\epsilon A_{ki} + \delta_{ki})} \\ &= A_{ji} - \delta_{ij} \sum_k A_{ki}. \end{aligned}$$

The theorem follows obviously. □

Next, we consider the similar limit in continuous states case.

Theorem 4.6 (Continuum Limit of the Backward Process). *For transition kernel $p_\epsilon(\mathbf{x}, \mathbf{y})$ and operator $\mathcal{P}_\epsilon f(\mathbf{x}) = \int p_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}$, define the corresponding backward counterparts*

$$\tilde{p}_\epsilon(\mathbf{x}, \mathbf{y}) = \frac{p_\epsilon(\mathbf{y}, \mathbf{x})}{\int p_\epsilon(\mathbf{y}, \mathbf{x}) d\mathbf{y}}, \quad \tilde{\mathcal{P}}_\epsilon f(\mathbf{x}) = \int \tilde{p}_\epsilon(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}.$$

Then if \mathcal{L} is the limit infinitesimal generator of \mathcal{P}_ϵ which satisfies

$$\mathcal{L}f = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\mathcal{P}_\epsilon f - f),$$

we have the infinitesimal generator of $\tilde{\mathcal{P}}_\epsilon$

$$\tilde{\mathcal{L}}f = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\tilde{\mathcal{P}}_\epsilon f - f) = \mathcal{L}^* f - f \mathcal{L}^* \mathbf{1}, \tag{4.34}$$

where \mathcal{L}^* is the conjugate operator of \mathcal{L} , i.e. $(\mathcal{L}f, g) = (f, \mathcal{L}^*g)$.

Proof. Consider the operator

$$\mathcal{P}_\epsilon^* g(\mathbf{x}) = \int p_\epsilon(\mathbf{y}, \mathbf{x}) g(\mathbf{y}) d\mathbf{y}.$$

It is straightforward that \mathcal{P}_ϵ^* is the conjugate of \mathcal{P}_ϵ , i.e. $(\mathcal{P}_\epsilon f, g) = (f, \mathcal{P}_\epsilon^* g)$, and we have

$$(f, \mathcal{L}^* g) = (\mathcal{L} f, g) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\mathcal{P}_\epsilon f - f, g) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (f, \mathcal{P}_\epsilon^* g - g).$$

This implies

$$\mathcal{L}^* g = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\mathcal{P}_\epsilon^* g - g), \quad \mathcal{P}_\epsilon^* g = g + \epsilon \mathcal{L}^* g + o(\epsilon).$$

Besides, it is obvious that

$$\tilde{\mathcal{P}}_\epsilon f(\mathbf{x}) = \frac{\mathcal{P}_\epsilon^* f(\mathbf{x})}{\mathcal{P}_\epsilon^* 1(\mathbf{x})}.$$

So we have

$$\begin{aligned} \tilde{\mathcal{L}} f(\mathbf{x}) &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\tilde{\mathcal{P}}_\epsilon f(\mathbf{x}) - f(\mathbf{x})) \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \left(\frac{\mathcal{P}_\epsilon^* f(\mathbf{x})}{\mathcal{P}_\epsilon^* 1(\mathbf{x})} - f(\mathbf{x}) \right) \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \left(\frac{\epsilon \mathcal{L}^* f(\mathbf{x}) + f(\mathbf{x}) + o(\epsilon)}{\epsilon \mathcal{L}^* 1(\mathbf{x}) + 1 + o(\epsilon)} - f(\mathbf{x}) \right) \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \left(\frac{\epsilon \mathcal{L}^* f(\mathbf{x}) - \epsilon f(\mathbf{x}) \mathcal{L}^* 1 + o(\epsilon)}{\epsilon \mathcal{L}^* 1(\mathbf{x}) + 1 + o(\epsilon)} \right) \\ &= \mathcal{L}^* f(\mathbf{x}) - f(\mathbf{x}) \mathcal{L}^* 1(\mathbf{x}). \end{aligned}$$

The proof is done. □

Now we apply Theorem 4.6 to two most relevant cases.

- ODE case: $\mathcal{L} f(\mathbf{x}) = v(\mathbf{x}) \cdot \nabla f(\mathbf{x})$.

In this case, we have

$$\tilde{\mathcal{L}} f(\mathbf{x}) = \mathcal{L}^* f(\mathbf{x}) - f(\mathbf{x}) \mathcal{L}^* 1(\mathbf{x}) = -v(\mathbf{x}) \cdot \nabla f(\mathbf{x}).$$

This corresponds to the ODE

$$\frac{d\mathbf{x}}{dt} = -v(\mathbf{x}),$$

which is exactly the reversed time dynamics of (4.31).

- SDE case: $\mathcal{L}f(\mathbf{x}) = \mathbf{v} \cdot \nabla f(\mathbf{x}) + \Delta f(\mathbf{x})$.

In this case, we have

$$\tilde{\mathcal{L}}f(\mathbf{x}) = \mathcal{L}^*f(\mathbf{x}) - f(\mathbf{x})\mathcal{L}^*1(\mathbf{x}) = -\mathbf{v}(\mathbf{x}) \cdot \nabla f(\mathbf{x}) + \Delta f(\mathbf{x}).$$

This corresponds to the SDEs

$$d\mathbf{x}(t) = -\mathbf{v}(\mathbf{x})dt + \sqrt{2}d\mathbf{w}(t).$$

To apply the theorem to the scRNA-seq data with transition rules (4.10) and (4.24), we need to take into account the data distribution $q(\mathbf{x})$. Define

$$\mathcal{P}_{q,\epsilon}f(\mathbf{x}) = \frac{\mathcal{P}_\epsilon(fq)(\mathbf{x})}{\mathcal{P}_\epsilon q(\mathbf{x})} = \frac{\int p_\epsilon(\mathbf{x}, \mathbf{y})f(\mathbf{y})q(\mathbf{y})d\mathbf{y}}{\int p_\epsilon(\mathbf{x}, \mathbf{y})q(\mathbf{y})d\mathbf{y}}.$$

It is not difficult to show that

$$\mathcal{L}_qf(\mathbf{x}) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\mathcal{P}_{q,\epsilon}f(\mathbf{x}) - f(\mathbf{x})) = \frac{\mathcal{L}(fq)(\mathbf{x}) - f(\mathbf{x})\mathcal{L}q(\mathbf{x})}{q(\mathbf{x})}.$$

Correspondingly, for

$$\tilde{\mathcal{P}}_{q,\epsilon}f(\mathbf{x}) = \frac{\tilde{\mathcal{P}}_\epsilon(fq)(\mathbf{x})}{\tilde{\mathcal{P}}_\epsilon q(\mathbf{x})},$$

we have

$$\begin{aligned} \tilde{\mathcal{L}}_qf(\mathbf{x}) &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} (\tilde{\mathcal{P}}_{q,\epsilon}f(\mathbf{x}) - f(\mathbf{x})) = \frac{\tilde{\mathcal{L}}(fq)(\mathbf{x}) - f(\mathbf{x})\tilde{\mathcal{L}}q(\mathbf{x})}{q(\mathbf{x})} \\ &= \frac{\mathcal{L}^*(fq)(\mathbf{x}) - f(\mathbf{x})\mathcal{L}^*q(\mathbf{x})}{q(\mathbf{x})}, \end{aligned} \quad (4.35)$$

where we utilized the conclusion in Theorem 4.6.

From Theorem 4.1 we know that for the cosine scheme with Gaussian diffusion kernel, $\mathcal{L}f = \hat{\mathbf{v}} \cdot \nabla f$ for appropriate $h(\cdot)$ and $g(\cdot)$. Then its conjugate operator $\mathcal{L}^*f = -\nabla \cdot (\hat{\mathbf{v}}f)$. Therefore, we have

$$\tilde{\mathcal{L}}_qf = -\hat{\mathbf{v}} \cdot \nabla f = -\mathcal{L}f.$$

Similarly, for correlation scheme, from Theorem 4.2 we also have

$$\tilde{\mathcal{L}}_qf = -\mathcal{L}f.$$

For the inner-product scheme, from Lemma 4.3 we know that

$$\mathcal{L}f = \frac{m_2}{2m_0g(0)} \mathcal{A}f = \frac{m_2}{2m_0g(0)} (g''(0)\|\mathbf{v}\|^2f + 2g'(0)\mathbf{v} \cdot \nabla f + g(0)\Delta f).$$

Then its conjugate operator is

$$\mathcal{L}^* f = \frac{m_2}{2m_0 g(0)} \mathcal{A}^* f = \frac{m_2}{2m_0 g(0)} (g''(0) \|v\|^2 f - 2g'(0) \nabla \cdot (vf) + g(0) \Delta f).$$

Similar to Theorem 4.3, simply replacing \mathcal{A} with its conjugate \mathcal{A}^* in the proof, we can show that

$$\tilde{\mathcal{L}}_\alpha f = \frac{m_2}{2m_0} \Delta f + \frac{m_2}{m_0} \left((1-\alpha) \frac{\nabla q}{q} - \frac{g'(0)}{g(0)} v \right) \cdot \nabla f,$$

which is similar as the operator \mathcal{L}_α (4.25) except reversing the direction of velocity v .

Overall, the above derivations show that for all of cosine, correlation and inner-product schemes, if we replace the transition kernel with its backward form, their continuum limit will follow the ODE or SDE dynamics, by reversing the direction of the RNA velocity v . Similar results also hold for kNN-cosine or kNN-correlation (or Gaussian-kNN-cosine/correlation) schemes with similar derivations. This gives the rationale of the identification of root and ending cells through backward and forward transition rules, respectively.

4.4 Finding development routes: Transition path theory

With the root and ending cells detected by RNA velocity, the next question is to ask how the cell state evolves along the development trajectories that connect the starting and target cell fates. Unlike the conventional picture of trajectory inference (such as pseudotimes), the dynamics revealed by RNA velocity might be more complex, because of local fluctuations, rotations and oscillations, as well as multiple sources and sinks along the trajectory. Except for calculating the most probable transition paths in the continuous set-up [41], the majority of existing tools opt to visualize the trajectories with local velocity arrows or connecting streamlines in the reduced-dimensional space, where a more quantified and global description of development path is needed.

We reason that the transition path theory [16] might be a good candidate, which has been established for general Markov process such as diffusion [15], jump [35] and Markov chains [36], and yielded fruitful applications in molecular dynamics and chemical reactions [5, 34]. Our proposed method to find development routes can be understood as a discrete, data-driven version of the continuous approach described by [41]. Below we only focus on the theoretical aspect of our proposal, whose algorithmic details will be discussed in the continued work [30]. We will mainly follow [34] for the illustration of the transition path theory, and ignore most proofs of the theorems which can be referred to [34] for details.

4.4.1 Coarse-graining of transition dynamics

The rapid growth of scRNA-seq data size poses computational challenges to the downstream analysis. Therefore we propose an optional step here to first coarse-grain the transition dynamics on the scale of clusters instead of single-cells to reduce the computational complexity.

Definition 4.1 (Coarse-graining of Markov Chain). *Given an ergodic, microscopic Markov chain $\{x_t\}$ on the state space S with transition probability matrix $P = \{p(x,y)\}$ and a partition $S = \cup_{k=1}^K S_k$, the coarse-graining of $\{x_t\}$ is defined as a Markov chain $\{X_t\}$ on the state space $\{S_k\}_{k=1}^K$ with transition probability matrix*

$$T_{ij} = \lim_{N \rightarrow \infty} \frac{\sum_{t=1}^N \mathbf{1}_{\{x_t \in S_i, x_{t+1} \in S_j\}}}{\sum_{t=1}^N \mathbf{1}_{\{x_t \in S_i\}}}, \tag{4.36}$$

where the limit is taken in the almost sure sense, and $\mathbf{1}_{\{\cdot\}}$ is the indicator function with $\mathbf{1}_{\{\text{exp}\}} = 1$ if the logical variable $\text{exp} = \text{TRUE}$, and 0 otherwise.

We remark that the naive upscaling of the microscopic Markov chain $\{x_t\}$ by considering the induced transition X_t on the coarse-grained space at each step is not valid since X_t defined in this way is not necessarily Markovian. This is related to the well-known lumpability concept in Markov chain theory [14, 27]. Here we take alternative viewpoint by defining the coarse-grained chain through the time average limit instead of the single step transition.

The partition of the state space can be achieved by the clustering of cells, or by the simultaneous reduction of multi-scale dynamics [14, 68]. The coarse-grained transition probability matrix can be indeed calculated analytically instead of through numerical simulations:

Proposition 4.1. *The coarse-grained transition probability matrix defined by (4.36) can be expressed as*

$$T_{ij} = \frac{\sum_{x \in S_i} \sum_{y \in S_j} \pi(x) p(x,y)}{\sum_{x \in S_i} \pi(x)}.$$

Proof. Consider the stochastic process $y_t = (x_t, x_{t+1})$, which is indeed a Markov chain with stationary distribution $\tilde{\pi}(x,y) = \pi(x)p(x,y)$. Then from the ergodic theorem for y_t we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \mathbf{1}_{\{x_t \in S_i, x_{t+1} \in S_j\}} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \mathbf{1}_{\{y_t \in (S_i, S_j)\}} = \sum_{x \in S_i} \sum_{y \in S_j} \pi(x) p(x,y),$$

and for x_t we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \mathbf{1}_{\{x_t \in S_i\}} = \sum_{x \in S_i} \pi(x).$$

The proof is done. □

It is also straightforward to verify that the coarse-grained Markov chain X_t with $\{T_{ij}\}$ has the stationary distribution μ with $\mu_i = \sum_{x \in S_i} \pi(x)$.

For simplicity of notations to present the transition path, below we take the index of $\{X_t\}$ as integer set \mathbb{Z} , where $t=0$ corresponds to the interested time point, and the minus time points represent the trajectories prior to X_0 . With this setup, X is stationary and X_t obeys the invariant distribution always.

4.4.2 Defining transition paths and their probabilities/fluxes

From the forward and backward transition approach described in Section 4.3, we are able to identify the sets of root and ending clusters, and denote them as committor starting set A and target set B , respectively. Note that both A and B may contain several states, corresponding to the complex dynamics of multiple root or ending states and various connecting trajectories. One advantage of transition path theory indeed lies in the quantification of such dynamics.

To begin with the derivation of transition path theory, we first define the core concepts of in-transition times and transition paths as follows.

Definition 4.2 (In-Transition Times). *For a given path $\{X_t\}$, the in-transition times from set A to B are defined as the union of sets*

$$\mathcal{T} = \bigcup_{n \in \mathbb{Z}} \{t \in \mathbb{Z} \mid t_n^A \leq t \leq t_n^B\},$$

where t_n^A and t_n^B are the n th exit and entrance time of set A and B respectively such that

$$X_{t_n^A} \in A, \quad X_{t_n^B} \in B, \quad X_t \in (A \cup B)^c \quad \text{for } t_n^A < t < t_n^B.$$

Definition 4.3 (Transition Paths). *For a given path $\{X_t\}$, the n th transition path from A to B is*

$$P_n = \{X_k \mid t_n^A \leq k \leq t_n^B\}.$$

The set of all transition paths is defined as $\mathcal{P} = \bigcup_{n \in \mathbb{Z}} \{P_n\}$.

We are interested in quantifying the probability distribution of transition paths ensemble, which is defined as:

Definition 4.4 (Probability of Transition Paths). *The probability of observing transition path at state i is defined as*

$$m_i^{AB} := \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N \mathbf{1}_{\{X_t=i\}} \mathbf{1}_{\{t \in \mathcal{T}\}}. \quad (4.37)$$

Intuitively, m_i^{AB} describes the likelihood that the cell is on a transition path from A to B and bypassing state S_i . To compute m_i^{AB} , we need the notion of committor functions and their associated first entrance/last exit time.

Definition 4.5 (First Entrance and Last Exit Time). *Given the path $\{X_t\}$ of stationary process X , the first entrance time τ_A^+ into set A , and the last exit time τ_A^- from set A are defined as*

$$\begin{aligned} \tau_A^+ &:= \inf\{t > 0: X_t \in A\}, \\ \tau_A^- &:= \inf\{t > 0: X_t^R \in A\}. \end{aligned}$$

where $X_t^R := X_{-t}$ is the path of time-reversed process of X_t , and X_t^R has the transition probability matrix $T_{ij}^R = \frac{\mu_j T_{ji}}{\mu_i}$.

Definition 4.6 (Committor Function). *The forward and backward committor functions are defined as*

$$q_i^+ := \mathbb{P}_i(\tau_B^+ < \tau_A^+), \quad q_i^- := \mathbb{P}_i^R(\tau_B^- > \tau_A^-).$$

Here \mathbb{P}_i denotes the probability of the forward process X conditioned on $X_0 = i$ and \mathbb{P}_i^R the probability of the reversed process X^R conditioned on $X_0^R = i$.

From the definition, we can also interpret q_i^+ as the probability that the cell starting from cluster S_i first enters set B rather than set A , and q_i^- the probability that the cell arriving at cluster S_i came last from set A instead of B . The committor functions provide a natural soft clustering of the states by measuring the affinities with starting set A or target set B .

To compute the committor functions, we can derive the linear equations they satisfy.

Proposition 4.2. *The committor functions solve the following discrete Dirichlet problems*

$$\begin{cases} \sum_k T_{ik} q_k^+ = q_i^+, & i \notin A \cup B, \\ q_i^+ = 0, & i \in A, \\ q_i^+ = 1, & i \in B. \end{cases} \quad \begin{cases} \sum_k T_{ik}^R q_k^- = q_i^-, & i \notin A \cup B, \\ q_i^- = 1, & i \in A, \\ q_i^- = 0, & i \in B. \end{cases}$$

Proof. Define the Markov chain X_t^A with absorbing set A through the transition probability

$$T_{ij}^A = \begin{cases} T_{ij}, & i \notin A, \\ 1, & i = j \in A, \\ 0, & i \in A, i \neq j \end{cases}$$

and the hitting time $\tau_B^A := \inf\{t > 0: X_t^A \in B\}$ of X_t^A into set B . Then we have $q_i^+ = \mathbb{P}_i(\tau_B^+ < \tau_A^+) = \mathbb{P}_i(\tau_B^A < +\infty)$. From the Markov property of X_t^A , for $i \in (A \cup B)^c$ we have

$$\mathbb{P}_i(\tau_B^A < +\infty) = \sum_k \mathbb{P}(\tau_B^A < +\infty | X_1 = k) \mathbb{P}_i(X_1 = k),$$

which yields the equation for q_i^+ , since $\mathbb{P}(\tau_B^A = m | X_0 = k) = \mathbb{P}(\tau_B^A = m + 1 | X_1 = k)$ for $k \in (A \cup B)^c$. The equation for q_i^- is similar. The validation of remaining boundary conditions are straightforward from the definition of committor function. \square

With committor functions, we have the following representation of transition paths probability.

Proposition 4.3. *The probability of transition paths defined in (4.37) can be expressed as*

$$m_i^{AB} = \mu_i q_i^- q_i^+.$$

The intuition of the above expression is clear. To observe the transition paths at state i , we pick it with the stationary distribution μ_i , and require that the path last exit from set A and first enter the set B . This happens with the probability q_i^- and q_i^+ , respectively.

Remark 4.9. The proportion ρ^{AB} of the time that a cell spends on the transition paths from A to B has the form

$$\rho^{AB} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N \mathbf{1}_{\{t \in \mathcal{S}\}} = \sum_{i=1}^K \mu_i q_i^- q_i^+.$$

Similarly, we can define the probability flux of transition paths, which is important to the detection of development routes discussed below.

Definition 4.7 (Probability Flux of Transition Paths).

$$f_{ij}^{AB} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \left(\mathbf{1}_{\{X_t=i, X_{t+1}=j\}} \cdot \sum_n \mathbf{1}_{\{t_n^A \leq t < t+1 \leq t_n^B\}} \right).$$

Roughly it tells the proportion of cells that are on a transition path from A to B and moving directly from S_i to S_j . We can also write f_{ij} in terms of committor functions, which also serves as the numerical strategy for computation.

Proposition 4.4. *The probability flux of transition paths can be expressed as*

$$f_{ij}^{AB} = \mu_i T_{ij} q_i^- q_j^+.$$

4.4.3 Finding development routes via transition paths flux

Trajectory inference aims to indicate how the state of cells evolve in a step-wise way. We therefore define the concept of development routes to illustrate this physical picture.

Definition 4.8 (Development Route). *A development route $\omega_{dr} = (i_0, i_1, \dots, i_n)$ from set A to B is a path connecting A and B without self-interactions (loops) such that*

$$i_0 \in A, i_n \in B, i_j \in (A \cup B)^c \text{ and } i_j \neq i_k, \text{ for } 0 \leq j \neq k \leq n.$$

To quantify development routes of Markov chain X_t , we need to eliminate the effect of detours along transition paths, and therefore define the notion of effective current (or net flux) based on the probability flux of transition paths.

Definition 4.9. *The effective current of transition paths from state i to j during the transition from set A to B is defined as*

$$f_{ij}^+ = \max\{f_{ij}^{AB} - f_{ji}^{AB}, 0\}.$$

With the effective current, we can specify the capacity of each development route.

Definition 4.10 (Capacity and Bottleneck). *Given the development route $\omega_{dr} = (i_0, i_1, \dots, i_n)$ from set A to B , its capacity is defined as*

$$c(\omega_{dr}) = \min_{0 \leq k \leq n-1} f_{i_k i_{k+1}}^+,$$

and the bottleneck is defined as the corresponding edge with minimal effective current.

The underlying intuition of the definition can be indeed understood by an analogy with the stream in water pipes or traffic on the freeways. The capacity of water pipes or freeways, is limited by the narrowest point where the minimal amount of water stream or traffic can pass through. Similarly, the transition of cell state from A to B is constrained by the bottleneck on the development route.

For all the possible development routes, we can calculate their capacity and consider routes with larger capacity, which corresponds to the more dominant trajectories during the transition. The ranking of all development routes capacity can be done effectively using an iterative edge-removing strategy [35]. We leave the details of algorithmic implementation of transition path theory to find development routes based on RNA velocity in our continued work [60].

5 Conclusion

The introduction of RNA velocity allows the prediction of future states in single-cell RNA sequence (scRNA-seq) data and yields fruitful results to reveal the dynamics of actual development process [48], while several theoretical issues regarding the models and analysis of RNA velocity remain to be elucidated. In this paper, we have proposed a mathematical framework to investigate the modeling, inference and downstream analysis aspects of RNA velocity.

Here we presented both the deterministic and stochastic models of RNA velocity, and derived the analytical solutions for both models. Particularly, we provided the expression for the exact probability distribution at any time in stochastic model. With the introduced models and analytical solutions, we then revisited the algorithms to infer parameters in RNA velocity model, and proposed an EM algorithm for the newly-derived complete stochastic model through maximum likelihood estimation.

Next, we dedicated to the theoretical issues on downstream analysis, particularly focusing on recovering the dynamical system models from RNA velocity. We derived the continuum limits of various constructed cell-cell transition dynamics from RNA velocity, which depended on the choice of velocity kernels (cosine, correlation or inner-product) and diffusion kernels (Gaussian or kNN). Our analysis revealed that while the cosine scheme in velocity kernel uncovers the streamlines in the deterministic ODE dynamics of RNA velocity model, the inner-product scheme corresponds to the stochastic dynamics described by SDE, which can incorporate the transitions among meta-stable states. Meanwhile the correlation scheme is associated with the deterministic dynamics with altered velocity that might be potentially unwanted. Through the delicate analysis on kNN kernel, we also proved that the choice of kNN over Gaussian in the diffusion kernel did not affect the overall continuum limit except for the pre-factors.

Based on our analysis, we then validated the rationale to find root and ending cells of previously proposed “forward and backward diffusion” strategy [29], from the continuous dynamical system aspects. It is shown that the difference between forward and backward transitions in the continuum limit only lies in the reversal of velocity direction. Finally, we proposed a method to infer the development routes from RNA velocity-based transition rules, which was specialized to cope with the complex dynamics of multiple root/ending states and various connecting path. We demonstrated the derivation of the method from transition path theory for Markov process.

Compared with previous stochastic models of RNA velocity that focused on moment equations of the chemical master equation [3, 41], the analysis of our proposed stochastic model featured for the derivation of an exact solution for the probability evolution. Such probability expression is especially useful to derive most-likelihood estimators of parameters for the stochastic model.

Currently, there exist two major streams of research to infer and analyze the underlying dynamics of scRNA-seq data. The first class of *data-driven* approaches, which are well represented by pseudo-time inference methods [21, 40, 53] for snapshot data, seek to construct development trajectories from the intrinsic manifold representation of data. On the other hand, the second class of *model-based* approaches as exemplified by SCUBA [33], Pseudo-dynamics [17] and Waddington-OT [43], aim to connect the probability distribution of single-cell data with a continuous dynamical system model, and are widely applied in the analysis of time-series sequencing data. Weinreb et al. [57] have shown theoretically that the fundamental limitations of snapshot data make the data-driven methods incapable of revealing the complex, non-equilibrium dynamics accurately. Meanwhile the model-based proposals may encounter difficulties in solving the high-dimensional Fokker-Planck equations numerically and infer the large amount of parameters in the model. Here we provide the mathematical justifications that the discrete cell-cell transition dynamics constructed from RNA velocity, even for snapshot data, can indeed converge into the continuous dynamical system model (not necessarily equilibrium) via the large sample limit, and we also propose the method to dissect the complex trajectories of such dynamics with solid theoretical guarantee in the well-established transition path

theory.

In the future, we anticipate that the RNA velocity model can be further improved by taking the genetic interactions into account [13], and customizing for the time-series scRNA-seq data. Overall, our mathematical analysis of RNA velocity in this paper provides the starting point to develop further models and methods in a more rational and consistent way.

Acknowledgments

The authors thank an anonymous referee for introducing the references [7, 24], in which the analytical solution of the stochastic RNA velocity model was also obtained, but re-discovered in this paper. T. Li and Y. Wu is supported by the NSFC under grants Nos. 11421101 and 11825102, and Beijing Academy of Artificial Intelligence (BAAI).

Appendix: The kNN radius estimate

In this Appendix, we will prove a useful kNN radius estimate as $n \rightarrow \infty$, which is summarized in Theorems A.1 and A.2. Although some classical results have been obtained in the literature concerning the consistency of the kNN density estimate [12, 31], the asymptotic expansion of the kNN radius is seldom discussed. Our analysis is more direct and independent of the result considered in [52]. It is not sharp, but enough for our continuum limit analysis of RNA velocity kernels. We leave further delicate order estimate as a future work.

Consider a random vector \mathbf{X} with smooth density function $q(\mathbf{x})$. Given n data points, we can estimate the density function at \mathbf{x} through kNN approach

$$q_n(\mathbf{x}) = \frac{k_n/n}{V_d R_n^d(\mathbf{x})},$$

where V_d is the volume of the d -dimensional unit sphere, $R_n(\mathbf{x})$ is the kNN radius defined as the distance to the k_n th nearest neighbor of \mathbf{x} . Without loss of generality, we set $\mathbf{x} = \mathbf{0}$ and denote $R_n = R_n(\mathbf{0})$. By definition, R_n obeys the order statistics with density

$$p_{k_n}(r) = \frac{n!}{(n-k_n)!(k_n-1)!} p(r) F^{k_n-1}(r) (1-F(r))^{n-k_n},$$

where $p(r)$ is the density of the radius $R = \|\mathbf{X}\|$:

$$p(r) = \int_{\|\mathbf{x}\|=r} q(\mathbf{x}) dS_{\mathbf{x}}, \quad r \geq 0 \tag{A.1}$$

and F is its distribution function $F(y) = \int_0^y p(r) dr$.

Below we will first estimate the expectation of R_n to achieve an intuition about its scale. We assume $q(\mathbf{0}) > 0$ all through the analysis.

Lemma A.1. *Under the condition $k_n \rightarrow \infty$, $k_n/n \rightarrow 0$ and the assumption*

$$\mathbb{E}\|\mathbf{X}\| = \int_0^\infty rp(r)dr = \int_0^1 F^{-1}(u)du < \infty,$$

we have

$$\bar{R}_n = \mathbb{E}R_n = c_d^{-\frac{1}{d}} \left(\frac{k_n}{n}\right)^{\frac{1}{d}} \left[1 + \mathcal{O}\left(\frac{k_n}{n}\right) + \mathcal{O}\left(\frac{k_n}{n}\right)^{\frac{2}{d}} + \mathcal{O}(k_n^{-1})\right], \quad (\text{A.2})$$

where $c_d = V_d q(\mathbf{0})$.

Proof. For any $0 < \epsilon \leq 1$, define

$$I_\epsilon(n) = \int_0^\epsilon F^{-1}(u)u^{k_n-1}(1-u)^{n-k_n}du, \quad I(n) = I_1(n).$$

Then we have

$$\begin{aligned} I(n) - I_\epsilon(n) &= \int_\epsilon^1 F^{-1}(u)u^{k_n-1}(1-u)^{n-k_n}du \\ &\leq \int_\epsilon^1 F^{-1}(u)(1-\epsilon)^{n-k_n}du \\ &\leq (1-\epsilon)^{n-k_n} \mathbb{E}\|\mathbf{X}\| = \mathcal{O}(\eta_\epsilon^n), \end{aligned}$$

where η_ϵ is a generic constant belongs to $(0,1)$, which can be taken as $\eta_\epsilon = (1-\epsilon)^{\frac{1}{2}}$ in the current step as long as $k_n \leq n/2$.

Now let us consider $I_\epsilon(n)$. We need to estimate the order of $F^{-1}(u)$ when u is small. For small $\delta > 0$, we have

$$\begin{aligned} F(\delta) &= \int_{\|\mathbf{x}\| \leq \delta} q(\mathbf{0}) + \mathbf{x}^T \nabla q(\mathbf{0}) + \frac{1}{2} \mathbf{x}^T \nabla^2 q(\mathbf{0}) \mathbf{x} + \mathcal{O}(\|\mathbf{x}\|^3) d\mathbf{x} \\ &= \delta^d V_d q(\mathbf{0}) + \frac{V_d \Delta q(\mathbf{0})}{2(d+2)} \delta^{d+2} + \mathcal{O}(\delta^{d+3}) \\ &= \delta^d (c_d + a_d \delta^2 + \mathcal{O}(\delta^3)) = \delta^d (c_d + \mathcal{O}(\delta^2)). \end{aligned} \quad (\text{A.3})$$

Denote $G(u) = F^{-1}(u)$, then $G(\delta^d(c_d + \mathcal{O}(\delta^2))) = \delta$. Let $u = \delta^d(c_d + \mathcal{O}(\delta^2))$. We get $\delta = \mathcal{O}(u^{1/d})$, and

$$\begin{aligned} G(u) = \delta &= \left(\frac{u}{c_d + \mathcal{O}(\delta^2)}\right)^{\frac{1}{d}} = \left(\frac{u}{c_d}\right)^{\frac{1}{d}} \left(1 - \frac{\mathcal{O}(\delta^2)}{c_d + \mathcal{O}(\delta^2)}\right)^{\frac{1}{d}} \\ &= \left(\frac{u}{c_d}\right)^{\frac{1}{d}} (1 + \mathcal{O}(\delta^2)) = \left(\frac{u}{c_d}\right)^{\frac{1}{d}} + \mathcal{O}(u^{\frac{3}{d}}). \end{aligned}$$

Therefore we obtain

$$\begin{aligned}
 I_\epsilon(n) &= \int_0^\epsilon G(u)u^{k_n-1}(1-u)^{n-k_n}du \\
 &= \int_0^\epsilon \left(\left(\frac{u}{c_d} \right)^{\frac{1}{d}} + \mathcal{O}(u^{\frac{3}{d}}) \right) u^{k_n-1}(1-u)^{n-k_n}du \\
 &= \int_0^1 \left(\left(\frac{u}{c_d} \right)^{\frac{1}{d}} + \mathcal{O}(u^{\frac{3}{d}}) \right) u^{k_n-1}(1-u)^{n-k_n}du + \mathcal{O}(\eta_\epsilon^n) \\
 &= c_d^{-\frac{1}{d}} B\left(k_n + \frac{1}{d}, n - k_n + 1\right) + \mathcal{O}\left(B\left(k_n + \frac{3}{d}, n - k_n + 1\right)\right) + \mathcal{O}(\eta_\epsilon^n),
 \end{aligned}$$

where $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ is the beta function. Utilizing the asymptotics

$$B(x, y) = \frac{\sqrt{2\pi}x^{x-\frac{1}{2}}y^{y-\frac{1}{2}}}{(x+y)^{x+y-\frac{1}{2}}} \left(1 + \mathcal{O}(x^{-1}) + \mathcal{O}(y^{-1})\right), \quad x, y \gg 1,$$

we obtain

$$\begin{aligned}
 B\left(k_n + \frac{1}{d}, n - k_n + 1\right) &= \Omega\left(\frac{1}{\sqrt{k_n}} \left(\frac{k_n}{en}\right)^{k_n + \frac{1}{d}}\right) \gg \mathcal{O}(\eta_\epsilon^n), \\
 B\left(k_n + \frac{3}{d}, n - k_n + 1\right) &\sim \left(\frac{k_n}{n}\right)^{\frac{2}{d}} B\left(k_n + \frac{1}{d}, n - k_n + 1\right) \gg \mathcal{O}(\eta_\epsilon^n),
 \end{aligned}$$

where $\Omega(\cdot)$ is the asymptotic lower bound function and we have used the condition $k_n/n \rightarrow 0$. Therefore, we get

$$I(n) = c_d^{-\frac{1}{d}} B\left(k_n + \frac{1}{d}, n - k_n + 1\right) \left(1 + \mathcal{O}\left(\frac{k_n}{n}\right)^{\frac{2}{d}}\right). \tag{A.4}$$

Then the expected radius

$$\begin{aligned}
 \bar{R}_n = \mathbb{E}R_n &= \int_0^\infty r p_{k_n}(r) dr \\
 &= \int_0^\infty r \frac{n!}{(n-k_n)!(k_n-1)!} p(r) F^{k_n-1}(r) (1-F(r))^{n-k_n} dr \\
 &= C_n \int_0^\infty r F^{k_n-1}(r) (1-F(r))^{n-k_n} dF(r) \\
 &= C_n \int_0^1 G(u) u^{k_n-1} (1-u)^{n-k_n} du = C_n I(n),
 \end{aligned}$$

where $C_n = n!/(n-k_n)!(k_n-1)!$. Using Stirling's formula and (A.4) we obtain (A.2) immediately. \square

Lemma A.1 suggests that the kNN radius $R_n(\mathbf{x})$ has the form $R_n(\mathbf{x}) = h_n r_n(\mathbf{x})$ with the scale

$$h_n = (k_n / (nV_d))^{1/d} \rightarrow 0$$

and the location dependent bandwidth

$$r_n(\mathbf{x}) = q(\mathbf{x})^{-\frac{1}{d}} + \mathcal{O}(h_n^d) + \mathcal{O}(h_n^2) + \mathcal{O}(k_n^{-1}).$$

Below we first establish the convergence in leading order in almost sure sense.

Theorem A.1. *If $k_n/n \rightarrow 0$ and $k_n/\ln n \rightarrow \infty$, then*

$$r_n(\mathbf{x}) \rightarrow q(\mathbf{x})^{-\frac{1}{d}} \quad \text{almost surely.}$$

Proof. Set $\mathbf{x} = \mathbf{0}$ without loss of generality. According to the Borel-Cantelli lemma, we only need to prove that for any small $\epsilon > 0$,

$$\sum_n \mathbb{P}(|r_n - q(\mathbf{0})^{-\frac{1}{d}}| > \epsilon) < +\infty. \quad (\text{A.5})$$

Denote $c = q(\mathbf{0})^{-\frac{1}{d}}$. We have

$$\begin{aligned} \mathbb{P}(|r_n - c| > \epsilon) &= \int_0^{c-\epsilon} C_n h_n p(rh_n) F^{k_n-1}(rh_n) (1 - F(rh_n))^{n-k_n} dr \\ &\quad + \int_{c+\epsilon}^{\infty} C_n h_n p(rh_n) F^{k_n-1}(rh_n) (1 - F(rh_n))^{n-k_n} dr \\ &:= P_\epsilon^- + P_\epsilon^+. \end{aligned}$$

From (A.3), we have the asymptotics

$$\begin{aligned} F^{k_n-1}(rh_n) &= \left((rh_n)^d (c_d + a_d r^2 h_n^2 + \mathcal{O}(r^3 h_n^3)) \right)^{k_n-1} \\ &= c_d^{k_n-1} (rh_n)^{d(k_n-1)} \left(1 + \frac{a_d}{c_d} r^2 h_n^2 + \mathcal{O}(r^3 h_n^3) \right)^{k_n-1} \\ &= c_d^{k_n-1} (rh_n)^{d(k_n-1)} \exp(r^2 o(k_n)) \end{aligned} \quad (\text{A.6})$$

and

$$\begin{aligned} (1 - F(rh_n))^{n-k_n} &= \left(1 - r^d h_n^d (c_d + a_d r^2 h_n^2 + \mathcal{O}(r^3 h_n^3)) \right)^{n-k_n} \\ &= \exp \left((n - k_n) \ln \left(1 - r^d h_n^d (c_d + a_d r^2 h_n^2 + \mathcal{O}(r^3 h_n^3)) \right) \right) \\ &= \exp \left((n - k_n) \ln \left(1 - \frac{r^d c_d k_n}{nV_d} \left(1 + \frac{a_d}{c_d} r^2 h_n^2 + \mathcal{O}(r^3 h_n^3) \right) \right) \right) \\ &= \exp \left(-\frac{r^d c_d k_n}{V_d} + (r^{d+2} + r^d) o(k_n) \right). \end{aligned} \quad (\text{A.7})$$

Define $g(x) = x^{k_n-1}(1-x)^{n-k_n}$. Then $(k_n-1)/(n-1)$ is the unique maxima of g , and g is increasing when $0 < x < (k_n-1)/(n-1)$ and decreasing when $(k_n-1)/(n-1) < x < 1$. Utilize (A.3), we can easily show that

$$F((c+\epsilon)h_n) > \frac{k_n-1}{n-1} \quad \text{and} \quad F((c-\epsilon)h_n) < \frac{k_n-1}{n-1}$$

for small ϵ and big enough n . Then using (A.6) and (A.7) and Stirling's formula, we obtain

$$\begin{aligned} P_\epsilon^+ &\leq \int_{c+\epsilon}^\infty C_n h_n p(rh_n) g(F((c+\epsilon)h_n)) dr \\ &\leq C_n g(F((c+\epsilon)h_n)) \\ &= C_n c_d^{k_n-1} (c+\epsilon)^{d(k_n-1)} h_n^{d(k_n-1)} \exp\left(-\frac{(c+\epsilon)^d c_d k_n}{V_d} + o(k_n)\right) \\ &= \mathcal{O}\left(\frac{n}{\sqrt{k_n}} \exp\left(k_n(f(c+\epsilon) - f(c) + o(1))\right)\right). \end{aligned}$$

where

$$f(r) = d \ln r - q(\mathbf{0}) r^d.$$

The function f is increasing when $r < c$ and decreasing when $r > c$.

With similar estimate for P_ϵ^- , we obtain

$$\mathbb{P}(|r_n - c| > \epsilon) = \mathcal{O}\left(\frac{n}{\sqrt{k_n}} \eta_\epsilon^{k_n}\right),$$

which is $o(n^{-2})$ under the condition $k_n/\ln n \rightarrow \infty$. Then (A.5) follows and the proof is done. □

Theorem A.2. *If $k_n/n \rightarrow 0$ and $k_n/n^\alpha \rightarrow \infty$ for some $\alpha > 2/(d+2)$, then for $d > 1$, we have*

$$\frac{r_n - q(\mathbf{0})^{-\frac{1}{d}}}{h_n} \xrightarrow{a.s.} 0, \quad \text{i.e.} \quad r_n = c + o(h_n) \quad \text{almost surely.}$$

Proof. Similar to Theorem A.1, now we consider

$$\begin{aligned} P_{\epsilon h_n} &= \int_0^{c-\epsilon h_n} + \int_{c+\epsilon h_n}^\infty C_n h_n p(rh_n) F^{k_n-1}(rh_n) (1-F(rh_n))^{n-k_n} dr \\ &:= C_{n,k_n,d} (I_{\epsilon h_n}^- + I_{\epsilon h_n}^+), \end{aligned}$$

where $C_{n,k_n,d} = C_n d c_d^{k_n} h_n^{d k_n}$ which has the asymptotics

$$C_{n,k_n,d} = d(q(0)e)^{k_n} \sqrt{\frac{k_n}{2\pi}} \exp\left(-\frac{k_n^2}{2n} + o\left(\frac{k_n^2}{n}\right)\right). \tag{A.8}$$

According to the proof of Theorem A.1, for any $\delta > 0$ and n big enough, we have

$$P_\delta \ll P_{\delta/2} \leq P_{\epsilon h_n}.$$

Therefore, to estimate $P_{\epsilon h_n}$, we only need to estimate $P_{\epsilon h_n} - P_\delta$. In what follows we will mainly use the Laplace's method to estimate $I_{\epsilon h_n}^+$ since the estimation of $I_{\epsilon h_n}^-$ is similar. First we have

$$\begin{aligned} p(rh_n) &= S_d(rh_n)^{d-1}q(\mathbf{0}) + S_d(rh_n)^{d+1}\frac{\Delta q(\mathbf{0})}{2d} + \mathcal{O}((rh_n)^{d+2}) \\ &= dc_d(rh_n)^{d-1} + \frac{1}{2}V_d\Delta q(\mathbf{0})(rh_n)^{d+1} + \mathcal{O}((rh_n)^{d+2}), \end{aligned} \quad (\text{A.9})$$

where S_d is the surface area of the d -dimensional unit sphere. Utilizing (A.6), (A.7) and (A.9), we get

$$\begin{aligned} I_{\epsilon h_n}^+ - I_\delta^+ &= \int_{c+\epsilon h_n}^{c+\delta} r^{dk_n-1} \exp\left(-\frac{r^d c_d k_n}{V_d} + f_1(r)k_n h_n^2 + f_2(r)\frac{k_n^2}{n}\right. \\ &\quad \left.+ o(k_n h_n^2) + o\left(\frac{k_n^2}{n}\right)\right) dr, \end{aligned}$$

where

$$f_1(r) = \frac{a_d}{c_d} r^2 \left(1 - \frac{r^d c_d}{V_d}\right), \quad f_2(r) = \frac{r^d c_d}{V_d} - \frac{r^{2d} c_d^2}{2V_d^2}.$$

We can verify that $f_1(c) = 0$ and $f_2(c) = 1/2$. By the arbitrary smallness of δ , we have

$$I_{\epsilon h_n}^+ - I_\delta^+ \sim \exp\left(o(k_n h_n^2) + \left(\frac{1}{2} + o(1)\right)\frac{k_n^2}{n}\right) \int_{c+\epsilon h_n}^{c+\delta} \frac{1}{r} \exp(k_n f(r)) dr.$$

Denote

$$J_\delta^+ = \int_{c+\epsilon h_n}^{c+\delta} \frac{1}{r} \exp(k_n f(r)) dr.$$

For $c + \epsilon h_n < r < c + \delta$, there exists $\bar{\xi} > 0$ such that

$$\left| f(r) - f(c + \epsilon h_n) - \frac{(r-c)^2 - \epsilon^2 h_n^2}{2} f''(c) \right| \leq \bar{\xi} \left((r-c)^2 - \epsilon^2 h_n^2 \right).$$

We have

$$\begin{aligned} J_\delta^+ &\geq \frac{\exp(k_n f(c + \epsilon h_n))}{c + \delta} \int_{c+\epsilon h_n}^{c+\delta} \exp\left(\left(\frac{f''(c)}{2} - \bar{\xi}\right)k_n((r-c)^2 - \epsilon^2 h_n^2)\right) dr \\ &\sim \frac{\exp(k_n f(c + \epsilon h_n))}{c + \delta} \int_{c+\epsilon h_n}^\infty \exp\left(\left(\frac{f''(c)}{2} - \bar{\xi}\right)k_n((r-c)^2 - \epsilon^2 h_n^2)\right) dr \\ &= \frac{h_n \exp(k_n f(c + \epsilon h_n))}{c + \delta} \int_\epsilon^\infty \exp\left(\left(\frac{f''(c)}{2} - \bar{\xi}\right)k_n h_n^2 (r^2 - \epsilon^2)\right) dr \\ &\sim \frac{\exp(k_n f(c + \epsilon h_n))}{2\epsilon k_n h_n (c + \delta) |f''(c)/2 - \bar{\xi}|}. \end{aligned}$$

Similar upper bound can be obtained, so we obtain

$$J_{\delta}^{+} = \mathcal{O}\left(\frac{\exp(k_n f(c + \epsilon h_n))}{k_n h_n}\right) = \mathcal{O}\left(\frac{\exp\left(k_n\left(f(c) + f''(c)\frac{\epsilon^2}{2}h_n^2 + o(h_n^2)\right)\right)}{k_n h_n}\right).$$

Therefore,

$$I_{\epsilon h_n}^{+} - I_{\delta}^{+} = \mathcal{O}\left(\frac{1}{k_n h_n} \exp\left(k_n\left(f(c) + \frac{1}{2}\epsilon^2 h_n^2 f''(c) + o(h_n^2) + \left(\frac{1}{2} + o(1)\right)\frac{k_n}{n}\right)\right)\right).$$

With similar estimate for $I_{\epsilon h_n}^{-} - I_{\delta}^{-}$, we get

$$\begin{aligned} P_{\epsilon h_n} &\sim P_{\epsilon h_n} - P_{\delta} = C_{n,k_n,d} \left((I_{\epsilon h_n}^{+} - I_{\delta}^{+}) + (I_{\epsilon h_n}^{-} - I_{\delta}^{-}) \right) \\ &= \mathcal{O}\left(\frac{1}{\sqrt{k_n} h_n} \exp\left(k_n\left(\left(\frac{1}{2}\epsilon^2 f''(c) + o(1)\right)h_n^2 + o\left(\frac{k_n}{n}\right)\right)\right)\right). \end{aligned}$$

If $d > 1$, then $h_n^2 = \mathcal{O}\left(\left(\frac{k_n}{n}\right)^{\frac{2}{d}}\right) > \mathcal{O}\left(\frac{k_n}{n}\right)$. Therefore

$$\mathbb{P}(|r_n - r| > \epsilon h_n) = \mathcal{O}\left(\frac{1}{\sqrt{k_n} h_n} \exp(-\eta_{\epsilon} k_n h_n^2)\right).$$

Under the condition that $k_n/n^{\alpha} \rightarrow \infty$ for some $\alpha > 2/(d+2)$, we have $k_n h_n^2/\ln n \rightarrow \infty$, so we get the desired estimate. \square

References

- [1] P. Ao. Potential in stochastic differential equations: novel construction. *J. Phys. A: Math. Gen.*, 37(3):L25, 2004.
- [2] C. Bender and S. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Springer-Verlag, New York, 1999.
- [3] V. Bergen, M. Lange, S. Peidli, F. A. Wolf, and F. J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.*, 2020.
- [4] T. Berry and T. Sauer. Local kernels and the geometric structure of data. *Appl. Comput. Harmon. Anal.*, 40(3):439–469, 2016.
- [5] G. R. Bowman, V. S. Pande, and F. Noé. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, volume 797. Springer Science & Business Media, 2013.
- [6] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, 36(5):411–420, 2018.
- [7] Z. Cao and R. Grima. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proc. Natl. Acad. Sci. USA*, 117(9):4682–4692, 2020.
- [8] W. Chung, H. H. Eum, H.-O. Lee, K.-M. Lee, H.-B. Lee, K.-T. Kim, H. S. Ryu, S. Kim, J. E. Lee, Y. H. Park, Z. Kan, W. Han, and W.-Y. Park. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.*, 8(1):1–12, 2017.

- [9] R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.
- [10] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, 2nd edition, 2006.
- [11] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39:1–38, 1977.
- [12] L. Devroye and T. Wagner. The strong uniform consistency of nearest neighbor density estimate. *Ann. Stat.*, 5:536–540, 1977.
- [13] J. Dong, P. Zhou, Y. Wu, W. Wang, Y. Chen, X. Zhou, H. Xie, Y. Gao, J. Lu, J. Yang, X. Zhang, L. Wen, W. Fu, T. Li, and F. Tang. Enhancing single-cell cellular state inference by incorporating molecular network features. *bioRxiv:699959*, 2019.
- [14] W. E, T. Li, and E. Vanden-Eijnden. Optimal partition and effective dynamics of complex networks. *Proc. Natl. Acad. Sci. USA*, 105:7907–7912, 2008.
- [15] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Stat. Phys.*, 123(3):503, 2006.
- [16] W. E and E. Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.*, 61:391–420, 2010.
- [17] D. S. Fischer, A. K. Fiedler, E. M. Kernfeld, R. M. Genga, A. Bastidas-Ponce, M. Bakhti, H. Lickert, J. Hasenauer, R. Maehr, and F. J. Theis. Inferring population dynamics from single-cell RNA-Sequencing time series data. *Nat. Biotechnol.*, 37(4):461–468, 2019.
- [18] H. Ge, M. Qian, and H. Qian. Stochastic theory of nonequilibrium steady states. Part II: Applications in chemical biophysics. *Phys. Rep.*, 510:87–118, 2012.
- [19] D. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. Academic Press, San Diego, 1992.
- [20] D. Gillespie. The chemical langevin equation. *J. Chem. Phys.*, 113(1):297–306, 2000.
- [21] L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, and F. J. Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods*, 13(10):845, 2016.
- [22] J. R. Heath, A. Ribas, and P. S. Mischel. Single-cell analysis tools for drug discovery and development. *Nat. Rev. Drug Discov.*, 15(3):204, 2016.
- [23] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.*, 94(12):128701, 2005.
- [24] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.*, 54(1):1–26, 2007.
- [25] D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, and I. Amit. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, 343(6172):776–779, 2014.
- [26] S. Jin, A. L. MacLean, T. Peng, and Q. Nie. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics*, 34(12):2077–2086, 2018.
- [27] J. Kemeny and J. Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976.
- [28] V. Y. Kiselev, K. Kirschner, M. T. Schaub, T. Andrews, A. Yiu, T. Chandra, K. N. Natarajan, W. Reik, M. Barahona, A. R. Green, and M. Hemberg. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods*, 14(5):483–486, 2017.
- [29] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastrioti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, and P. V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018.
- [30] T. Li, J. Shi, Y. Wu, and P. Zhou. On the mathematics of RNA velocity II: algorithmic aspects.

In preparation, 2020.

- [31] D. Loftsgaarden and C. Quesenberry. A nonparametric estimate of a multivariate density function. *Ann. Math. Stat.*, 36:1049–1051, 1965.
- [32] E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, and S. A. McCarroll. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.
- [33] E. Marco, R. L. Karp, G. Guo, P. Robson, A. H. Hart, L. Trippa, and G.-C. Yuan. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci. USA*, 111(52):E5643–E5650, 2014.
- [34] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, 125(8):084110, 2006.
- [35] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Transition path theory for Markov jump processes. *Multiscale Model. Simul.*, 7(3):1192–1219, 2009.
- [36] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA*, 106(45):19011–19016, 2009.
- [37] A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev, and B. E. Bernstein. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190):1396–1401, 2014.
- [38] S. Picelli, Å. K. Björklund, O. R. Faridani, S. Sagasser, G. Winberg, and R. Sandberg. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, 10(11):1096–1098, 2013.
- [39] Q. Qiu, P. Hu, X. Qiu, K. W. Govek, P. G. Cámara, and H. Wu. Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nat. Methods*, pages 1–11, 2020.
- [40] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, and C. Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*, 14(10):979, 2017.
- [41] X. Qiu, Y. Zhang, D. Yang, S. Hosseinzadeh, L. Wang, R. Yuan, S. Xu, Y. Ma, J. Replogle, S. Darmanis, J. Xing, and J. S. Weissman. Mapping vector field of single cells. *bioRxiv:696724*, 2019.
- [42] A.-E. Saliba, A. J. Westermann, S. A. Gorski, and J. Vogel. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.*, 42(14):8845–8860, 2014.
- [43] 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. *Cell*, 176(4):928–943, 2019.
- [44] Y. Sha, S. Wang, P. Zhou, and Q. Nie. Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. *Nucleic Acids Res.*, 2020.
- [45] V. Shahrezaei and P. Swain. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA*, 105:17256–17261, 2008.
- [46] J. Shi, T. Li, L. Chen, and K. Aihara. Quantifying pluripotency landscape of cell differentiation from scRNA-seq data by continuous birth-death process. *PLoS Comput. Biol.*, 15(11):e1007488, 2019.
- [47] J. Shi, A. E. Teschendorff, W. Chen, L. Chen, and T. Li. Quantifying Waddington’s epigenetic landscape: a comparison of single-cell potency measures. *Briefings Bioinf.*, 21(1):248–261, 2020.

- [48] V. Svensson and L. Pachter. RNA velocity: molecular kinetics from single-cell RNA-Seq. *Mol. Cell*, 72(1):7–9, 2018.
- [49] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*, 99(20):12795–12800, 2002.
- [50] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao, and M. A. Surani. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*, 6(5):377–382, 2009.
- [51] A. E. Teschendorff and T. Enver. Single-cell entropy for accurate estimation of differentiation potency from a cell’s transcriptome. *Nat. Commun.*, 8(1):1–15, 2017.
- [52] D. Ting, L. Huang, and M. Jordan. An analysis of the convergence of graph Laplacians. *Proc. 27th Int. Conf. Mach. Learn.*, 2010.
- [53] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, and J. L. Rinn. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.*, 32(4):381, 2014.
- [54] B. Treutlein, D. G. Brownfield, A. R. Wu, N. F. Neff, G. L. Mantalas, F. H. Espinoza, T. J. Desai, M. A. Krasnow, and S. R. Quake. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, 509(7500):371–375, 2014.
- [55] C. Waddington. *The Strategy of the Genes: a Discussion of Some Aspects of Theoretical Biology*. Allen & Unwin, 1957.
- [56] J. Wang, L. Xu, and E. Wang. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. USA*, 105(34):12271–12276, 2008.
- [57] C. Weinreb, S. Wolock, B. Tusi, M. Socolovsky, and A. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. USA*, 115:E2467–E2476, 2018.
- [58] A. R. Wu, N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, and S. R. Quake. Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods*, 11(1):41, 2014.
- [59] H. Wu, C. Wang, and S. Wu. Single-cell sequencing for drug discovery and drug development. *Curr. Top. Med. Chem.*, 17(15):1769–1777, 2017.
- [60] Y. Wu, P. Zhou, J. Shi, and T. Li. Inferring the RNA velocity based on the stochastic dynamics. In preparation, 2020.
- [61] J. Xing. Mapping between dissipative and Hamiltonian systems. *J. Phys. A: Math. Theor.*, 43(37):375003, 2010.
- [62] L. Zappia, B. Phipson, and A. Oshlack. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.*, 14(6):e1006245, 2018.
- [63] J. Zhang, Q. Nie, and T. Zhou. Revealing dynamic mechanisms of cell fate decisions from single-cell transcriptomic data. *Front. Genet.*, 10:1280, 2019.
- [64] X. Zhang, S. L. Marjani, Z. Hu, S. M. Weissman, X. Pan, and S. Wu. Single-cell sequencing for precise cancer research: progress and prospects. *Cancer Res.*, 76(6):1305–1312, 2016.
- [65] X. Zhang, H. Qian, and M. Qian. Stochastic theory of nonequilibrium steady states and its applications. Part I. *Phys. Rep.*, 510:1–86, 2012.
- [66] G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, and J. H. Bielas. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*,

8(1):1–12, 2017.

- [67] P. Zhou and T. Li. Construction of the landscape for multi-stable systems: potential landscape, quasi-potential, A-type integral and beyond. *J. Chem. Phys.*, 144:94109, 2016.
- [68] P. Zhou, S. Wang, T. Li, and Q. Nie. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. preprint, 2020.