bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Optimal transport analysis reveals trajectories in steady-state systems

Stephen Zhang, Anton Afanassiev, Laura Greenstreet, Tetsuya Matsumoto, Geoffrey Schiebinger*

Department of Mathematics, University of British Columbia, Vancouver, BC, Canada

**Abstract**

Understanding how cells change their identity and behaviour in living systems is an important question in many fields of biology. The problem of inferring cell trajectories from single-cell measurements has been a major topic in the single-cell analysis community, with different methods developed for equilibrium and non-equilibrium systems (e.g. haematopoeisis vs. embryonic development). We show that optimal transport analysis, a technique originally designed for analysing time-courses, may also be applied to infer cellular trajectories from a single snapshot of a population in equilibrium. Therefore optimal transport provides a unified approach to inferring trajectories, applicable to both stationary and non-stationary systems. Our method, StationaryOT, is mathematically motivated in a natural way from the hypothesis of a Waddington's epigenetic landscape. We implemented StationaryOT as a software package and demonstrate its efficacy when applied to simulated data as well as single-cell data from *Arabidopsis thaliana* root development.

# 1 Introduction

## 1.1 Background

Biological processes at the cellular level are driven by stochastic dynamics – cellular populations evolve through time, driven by regulation at the cellular and tissue level and intrinsic noise arising from molecular and thermal fluctuations. In the context of developmental biology, these processes have been classically described by Waddington's metaphor of an epigenetic landscape [39], in which differentiating cells can be thought of as evolving from regions of high differentiation potential into valleys corresponding to differentiated cell types. In the last decade, this metaphor has evolved to be much more quantitative [38, 29]. Modern high-throughput assays such as single-cell RNA sequencing (scRNA-seq) [15, 23], scATAC-seq [4] and CyTOF [24] now allow the molecular states of thousands of single cells to be profiled in a single experiment. With the ability to make these precision measurements of cell state, new challenges emerge in analysing these new types of high-dimensional data.

Single-cell measurements are destructive in nature, so the state of any individual cell cannot be observed at more than one instant. Therefore, information about the trajectories taken by cells over time is lost and must instead be inferred from data. A large collection of trajectory inference methods have been developed in recent years [38] to address this issue. These methods broadly fall into two classes [20]: (1) methods that deal with a single stationary snapshot observed from a cellular population at equilibrium [42, 37, 31], and (2) methods that deal with a time series of snapshots from an evolving population [29, 36, 20].

Time-series experiments are a natural approach for observing biological systems where cellular populations undergo dramatic, synchronous changes, such as in embryogenesis or stem-cell reprogramming [33, 43, 29, 40, 21]. Trajectory inference methods for time series data primarily seek to infer cellular transition events from snapshots of one timepoint to the next. On the other hand, development occurs continuously and asynchronously in many biological systems such as haematopoiesis and spermatogenesis. These systems maintain a stationary population profile across various cell types and can be thought of as being in dynamic equilibrium (i.e. steady state). Snapshots therefore capture cells from across the full progression of cell states from undifferentiated to fully differentiated cells. Trajectory inference for snapshots sampled from these steady-state systems seek to (a) infer the progression of cells in "developmental time" (commonly referred to as pseudotime) [37, 13], and (b) uncover bifurcation events or "cellular decisions" occurring in the differentiation process [42, 41].

To date, the methods developed for these two different paradigms have remained largely distinct, with time-course methods generally sharing little in common with methods designed for stationary systems.

*Corresponding author, ✉geoff@math.ubc.ca

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In this paper we show that optimal transport analysis, a technique originally designed for analysing time-courses [29], may also be applied to infer cellular trajectories from a single snapshot of a population in equilibrium. Therefore, optimal transport (OT) provides a unified approach to inferring trajectories, applicable to both stationary and non-stationary systems.

Our approach is theoretically justified when the trajectories are driven by a potential landscape, as in [41]. Moreover, it also allows for extensions to incorporate additional information such as estimates of the vector field [16, 1]. When such information is available, we can recover certain aspects of non-conservative dynamics such as oscillations.

## 1.2 Modelling assumptions

**Development as drift-diffusion with birth-death** We model cells as points in a space $\mathcal{X}$, which we take to be a representation of the space of possible cellular molecular states (for instance, in the case of scRNA-seq data, $\mathcal{X}$ represents the space of gene expression profiles). Typically, we will take $\mathcal{X} = \mathbb{R}^d$. We regard cells as evolving following a drift-diffusion process [41, 18, 3] described by the SDE

$$\mathrm{d}X_t = \mathbf{v}(X_t)\mathrm{d}t + \sigma\, \mathrm{d}B_t. \tag{1}$$

where $X_t \in \mathcal{X}$ is the state of a cell at time $t$, $\mathbf{v}$ is a vector field, the diffusivity $\sigma^2$ captures the noise level and $\mathrm{d}B_t$ denotes the increments of a Brownian motion. In addition, cells are subject to division and death events at exponential rates $\beta(x)$ and $\delta(x)$ respectively, which may vary in spatial location in $\mathcal{X}$. That is, in an infinitesimal time interval $\mathrm{d}t$, a cell $X_t$ may divide with probability $\beta(X_t)\,\mathrm{d}t$ or die with probability $\delta(X_t)\,\mathrm{d}t$.

**Population-level model** At the population level, the drift-diffusion process with birth and death can be described by a *population balance* partial differential equation (PDE) [41, 9]

$$\partial_t \rho(x,t) = -\nabla \cdot (\mathbf{v}(x)\rho(x,t)) + \frac{\sigma^2}{2}\nabla^2 \rho(x,t) + R(x)\rho(x,t), \tag{2}$$

Where $\rho(x,t)$ is a continuous population density, and $R$ is a spatially varying flux rate defined as $R(x) = \beta(x) - \delta(x)$ that captures creation and destruction of cells due to birth and death, as well as entry and exit from the system.

**Sources and sinks** In most developmental systems, cells eventually exit the observed system once they have fully differentiated. For instance, terminally differentiated immune cells leave the haematopoietic stem cell niche. This can be incorporated into the mathematical model by introducing a set of "sink regions" $\mathcal{X}_\varnothing \subset \mathcal{X}$ where cells are removed at some rate once they enter [41]. Mathematically, this can be treated in an identical way to cell death, and we incorporate it into the death rate $\delta(x)$ by setting $\delta(x) > 0$ whenever $x \in \mathcal{X}_\varnothing$. This can then be interpreted that in time $\mathrm{d}t$, a cell $X_t$ located in the sink region exits with probability $\delta(X_t)\,\mathrm{d}t$.

**Observation model** As we discussed in Section 1.1, many biological processes exist approximately in an equilibrium or steady state. In this setting, a snapshot at a single instant in time will capture all stages of cellular development in the system [41, 38], and relative proportions of various cell types remain unchanged over time. Mathematically at the population level, this assumption amounts to demanding that $\partial_t \rho(x,t) = 0$ in (2), that is, the population level cell density does not change. We will write $\rho_{\mathrm{eq}}(x)$ for this steady-state solution. Experimental observation of such a system is therefore equivalent to sampling a collection of $N$ cellular states from the population, i.e.

$$X_1, \ldots, X_N \sim \rho_{\mathrm{eq}}.$$

We may describe this finite sample as an empirical distribution $\hat{\rho}_{\mathrm{eq}}$ supported on the discrete space $\overline{\mathcal{X}} = \{X_i\}_{i=1}^N$,

$$\hat{\rho}_{\mathrm{eq}} = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}. \tag{3}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

### 1.3 Inference goal

**Laws on paths** The process introduced in Section 1.2 is a superposition of a birth-death process and a drift-diffusion process (1). The drift-diffusion component (1) of this process governs the state dynamics of individual cells, and can be related directly to the concept of a Waddington's potential landscape. Therefore, we seek to learn something about the drift-diffusion dynamics from observed snapshot data. In the framework of stochastic differential equations which we employ, the process (1) (equipped also with an initial condition) can be thought of as inducing a probability distribution over the space of possible trajectories of cells, which we take to be the set of continuous paths in $\mathcal{X}$ parameterised by time, i.e. $\Omega = C([0,1], \mathcal{X})$. As we also argue in [18], this *law on paths* is the natural object we seek to estimate since it directly encodes the trajectories that cells may follow. Furthermore, the law on paths can be obtained as the solution to a convex problem, so we do not have issues of multiple local minima which may be the case if we attempt to recover the drift field $\mathbf{v}$ or potential landscape $\Psi$ directly.

**Identifiability** For the sake of making inferences about the law on paths induced by (1), we must necessarily have estimates of the flux rate $R(x)$ and the noise level $\sigma^2$. As noted by [41], when only a single snapshot (i.e. $\hat{\rho}_{\text{eq}}$) is available, in general more than one drift field $\mathbf{v}$ can give rise to the same steady-state density profile $\rho_{\text{eq}}$. To ensure uniqueness of the solution, we must restrict to the case where the drift is given by the gradient of a scalar potential [41, 18], i.e. $\mathbf{v} = -\nabla\Psi$.

### 1.4 Related work

Weinreb et al. [41] previously investigated this problem and noted that the drift was identifiable only in gradient-driven systems. In addition, the authors presented population balance analysis (PBA), a methodological framework for estimating the potential $\Psi$ based on spectral graph theory. Although our approach and that of [41] share a problem formulation and may indeed perform similarly, we note that the theoretical foundations of the two approaches are fundamentally distinct – our method is based on solving a convex optimisation problem for the transition probabilities, whilst PBA solves a system of linear equations for the potential. As an optimisation-based method, optimal transport also allows for incorporation of additional information such as velocity estimates.

Optimal transportation (OT) theory is a mathematical area of study concerned with optimally coupling probability distributions [25] which has recently found diverse applications in statistics, machine learning and computational geometry. Optimal transport has been applied to the problem of tracking particle ensembles [12, 6], and to single-cell trajectory inference in the setting of time-series population snapshots in [29]. Subsequent work has extended both methodology and theory in this direction, e.g. [35, 18, 10, 27, 7]. However, these works focus on the setting where multiple snapshots are available over a series of time-points. We show in this work that optimal transport can be applied in a natural way to the case of a single stationary snapshot, further establishing optimal transport as a widely applicable and robust framework for single-cell trajectory inference.

## 2 Results

**Overview of results** To motivate the mathematical framework for our method, we will consider first the population-level setting of infinitely many cells (Section 2.1). We then reduce this to the discrete setting where we deal with finite samples drawn from the steady state population (Section 2.3). We name our method StationaryOT and implement it as a software package (see Software availability below the Discussion). In Sections 2.6 and 2.7 we apply the method to simulated datasets sampled from drift-diffusion processes in the setting of both potential-driven and non-conservative vector fields. In Section 2.8 we demonstrate an application to a stationary snapshot scRNA-seq dataset in *Arabidopsis thaliana* root tip development. Finally in Section 2.9 we discuss approaches for applying StationaryOT to very large datasets, showing that our method can scale to $1.1 \times 10^5$ cells with runtimes of $\sim$1 hour.

Throughout this paper, we consider using either entropy-regularised (Section 2.2) or quadratically-regularised optimal transport (Section 2.4) for the main step of the StationaryOT method. Although entropy-regularised optimal transport is the one that arises naturally from the theoretical motivation, we demonstrate in practice that using the quadratic regularisation generally leads to results that are more robust and interpretable, as well as being more computationally favourable thanks to sparsity of the recovered transition laws, without any substantial sacrifices to accuracy.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## 2.1 Methodology: population level

At the steady state of the process (2), the population density profile is constant, i.e. $\rho(\cdot, t) = \rho_{\text{eq}}$. However, at the microscopic level, individual cells $X_t$ continue to undergo drift-diffusion as described by (1), as well as birth-death. Thus, observation of population profiles in the stationary setting do not contain information about the dynamics of individual particles, unlike the non-stationary setting of time series measurements [29, 18].

Suppose that we are able to observe a cell $X_t$ from the stationary population at time $t$, and again at time $t + \Delta t$ (conditioned on not dividing, dying or exiting from the system in that time interval). Then the joint distribution $(X_t, X_{t+\Delta t})$ would capture information about all possible transitions in cell state over a time interval $\Delta t$. The system is at a steady state and the dynamics are Markov, so knowledge of the time-$\Delta t$ evolution of the system captures the full law on paths that results from the drift-diffusion component (1), at least at times $\{k\Delta t, k = 0, 1, \ldots\}$ by simply composing Markov transitions.

Since we may access densities but not track individual particles, we cannot measure the joint distribution $(X_t, X_{t+\Delta t})$ directly. Therefore, we seek to infer it from observation of a single snapshot $\rho_{\text{eq}}$ and information about the birth-death rates as well as noise level. In the underlying process both birth-death and drift-diffusion take place simultaneously, leading to complications in directly reasoning with probability laws. In order to simplify this, we approximate the evolution of the process by introducing an artificial separation of the effects of growth and transport, inspired by operator splitting methods from numerical analysis [14]. That is, we split the linear equation (2) into equations corresponding to growth and transport in the densities $\rho_G$ and $\rho_T$ respectively:

$$\frac{\partial \rho_G}{\partial t} = R(x)\rho_G(x, t), \tag{4}$$

$$\frac{\partial \rho_T}{\partial t} = -\nabla \cdot (\mathbf{v}(x)\rho_T(x, t)) + \frac{\sigma^2}{2}\nabla^2 \rho_T(x, t) \tag{5}$$

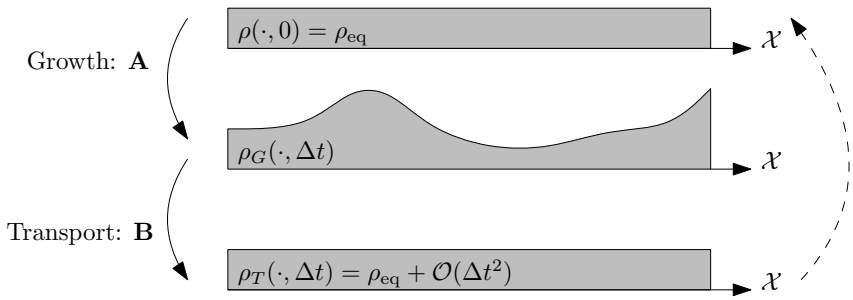
$$\text{where } \rho_G(\cdot, 0) = \rho_{\text{eq}}(\cdot) \text{ and } \rho_T(\cdot, 0) = \rho_G(\cdot, \Delta t). \tag{6}$$

Then $\rho_T(\cdot, \Delta t)$ is a splitting approximation of the true steady-state solution $\rho(\cdot, \Delta t) = \rho_{\text{eq}}(\cdot)$ of (2), and the two coincide in the limit $\Delta t \to 0$ with approximation error of order $\mathcal{O}(\Delta t^2)$ [14, Section 1.3], i.e.

$$\rho_T(x, \Delta t) = \rho(x, \Delta t) + \mathcal{O}(\Delta t^2),\ x \in \mathcal{X}. \tag{7}$$

We provide a conceptual illustration of this scheme in Figure 1. The solution of the growth step (4) can be determined to be exactly

$$\rho_G(x, \Delta t) = \rho(x, 0)e^{\Delta t R(x)} = \rho(x, 0)g(x)^{\Delta t}.$$




**A.** $\partial_t \rho_G = R(x)\rho_G(x, t)$, $\rho_G(\cdot, 0) = \rho(\cdot, 0)$
**B.** $\partial_t \rho_T = -\nabla \cdot (\mathbf{v}(x)\rho_T(x, t)) + \frac{1}{2}\sigma^2 \nabla^2 \rho_T(x, t)$, $\rho_T(\cdot, 0) = \rho_G(\cdot, \Delta t)$

Figure 1: Illustration of the splitting scheme for decomposing (2) into (4), (5).

It therefore remains for us to examine the effects due to transport in the equation (5). Since the overall system (2) is assumed to be at steady state, composing the effects of growth and transport should yield the initial density $\rho(\cdot, 0) = \rho_{\text{eq}}(\cdot)$ up to the $\mathcal{O}(\Delta t^2)$ error introduced by the splitting approximation. For brevity, let us denote $\mu_0 = \rho_T(\cdot, 0)$ and $\mu_1 = \rho_T(\cdot, \Delta t)$ to be the distributions before and after transport. Under the splitting approximation, our problem of estimating the joint law $(X_t, X_{t+\Delta t})$ conditional on no birth or death events amounts to finding an appropriate coupling $\gamma_{\Delta t}$ of $(\mu_0, \mu_1)$, i.e. a joint distribution $\gamma_{\Delta t}$ on $\mathcal{X}^2$ whose marginals agree with $\mu_0$ and $\mu_1$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## 2.2 Inference by optimal transport

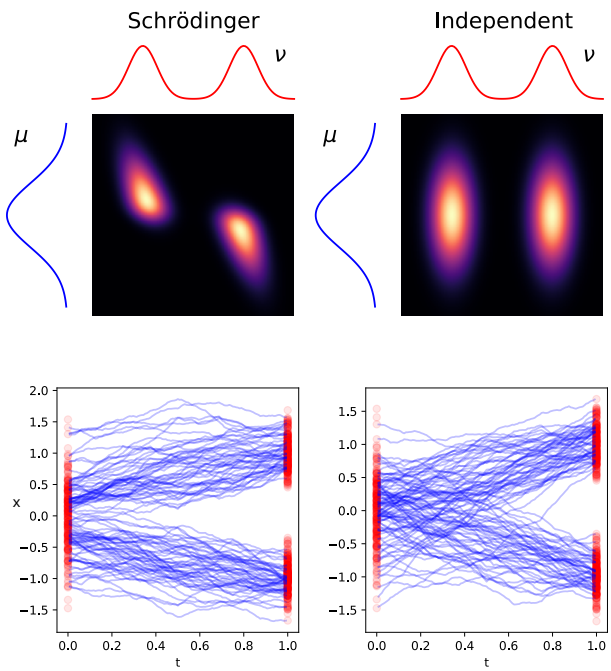



Figure 2: Examples of possible couplings of two prescribed distributions $(\mu, \nu)$

By the previous construction, we seek to couple the distributions $(\mu_0, \mu_1)$ in a way that approximates the "true" underlying transition law $(X_t, X_{t+\Delta t})$. To be concise, we write

$$\Pi(\mu_0, \mu_1) = \left\{ \pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{X}) : \int \pi(dx, \cdot) = \mu_1, \int \pi(\cdot, dy) = \mu_0 \right\}$$

to denote the set of possible couplings between $\mu_0$ and $\mu_1$. As we show visually in Figure 2, for a set of prescribed marginals there are in general many valid couplings. Indeed, for any $\mu, \nu$ we may always construct the independent coupling, $\mu \otimes \nu$. Therefore, additional assumptions on the nature of the process driving the evolution from $\mu_0$ to $\mu_1$ are needed if we desire a unique "best" coupling.

From the drift-diffusion splitting step (5), we know that the evolution from $\mu_0$ to $\mu_1$ is described by a drift-diffusion equation (with no source term). At the level of individual particles, this is equivalent to the stochastic differential equation (1). We note further that for $\Delta t$ small, the effect of the drift component is $\mathcal{O}(\Delta t)$ and is therefore drowned out by the effect of the diffusion component which is $\mathcal{O}(\sqrt{\Delta t})$. Thus, for small $\Delta t$, the setting which we approach is that of a diffusive evolution in time $\Delta t$ from $\mu_0$ to $\mu_1$, and the most likely coupling $\gamma_{\Delta t}$ is unique and is characterised by an entropy minimisation principle that is well known in the literature of optimal transport and large deviation theory [19]. Specifically, the optimal coupling $\gamma_{\Delta t}$ is the minimiser of the so-called *Schrödinger problem*:

$$\min_{\gamma_{\Delta t} \in \Pi(\mu_0, \mu_1)} \mathrm{H}(\gamma_{\Delta t} | K_{\sigma^2 \Delta t}). \tag{8}$$

In the above, $\mathrm{H}(\alpha|\beta) = \int \mathrm{d}\alpha \log\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\right)$ is the relative entropy between distributions, and $K_{\sigma^2 \Delta t}$ is the kernel

$$K_{\sigma^2 \Delta t}(x, y) = \exp\left(-\frac{1}{2\sigma^2 \Delta t} \|x - y\|^2\right),$$

corresponding to the time-$\Delta t$ evolution of a Brownian motion in $\mathcal{X}$ with diffusivity $\sigma^2$.

The problem (8) is also known in the optimal transport literature as entropy-regularised optimal transport [25], where the objective to be minimised is often written in the alternative form

$$\min_{\gamma_{\Delta t} \in \Pi(\mu_0, \mu_1)} \int C(x, y) \mathrm{d}\gamma_{\Delta t}(x, y) + \varepsilon \mathrm{H}(\gamma_{\Delta t} | \mathrm{Leb}) \tag{9}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $C(x, y) = \frac{1}{2}\|x-y\|^2$ is a quadratic cost function, $\varepsilon = \sigma^2\Delta t$ is the entropy regularisation parameter and Leb is the reference Lebesgue measure on $\mathcal{X}$. Written in this way, (9) can be understood as a least action principle, where the optimal $\gamma_{\Delta t}$ is roughly the one that minimises the expected action for moving mass from $\mu_0$ to $\mu_1$, if the action is proportional to the squared distance moved. This can be noted from Figure 2, where the coupling given by the solution of the Schrödinger problem results in trajectories that bifurcate simply. On the other hand for the other couplings, we get "crossing" of trajectories which result in a higher energetic cost. In the limiting case of vanishing noise where $\varepsilon \to 0$, the entropy-regularised optimal transport problem becomes what is known as the Monge-Kantorovich problem, or unregularised optimal transport [25].

We conclude that the coupling $\gamma_{\Delta t}$ recovered by solving the entropy minimisation problem (8) for couplings in $\Pi(\mu_0, \mu_1)$ is an approximation to the true evolution of the process (1), corresponding to the drift-diffusion step (5) of the splitting scheme.

## 2.3 Methodology: finite samples

**Formulation of the discrete problem** In practice, we have access to an empirical distribution $\hat{\rho}_{\text{eq}}$ (3) supported on the discrete set $\overline{\mathcal{X}}$ that can be thought of as approximating the true continuous density $\rho_{\text{eq}}$ discussed previously. We also assume for each observed cell $x_i$ that we have an estimate of the corresponding flux rate $\hat{R}_i = R(x_i) = \beta(x_i) - \delta(x_i)$. In a practical biological setting, cell states which are expected to divide or die should therefore have $\hat{R}_i > 0$ or $\hat{R}_i < 0$ respectively, and those states which do neither should have $\hat{R}_i = 0$. In addition to division and death, terminally differentiated cells expected to shortly exit the system may be regarded as representing sinks, and therefore assigned $\hat{R}_i < 0$. The numerical values for flux rates may be estimated from cell-cycle signatures [29] or prior biological knowledge [41].

In this discrete setting, the growth step (4) is local in space and thus its analogue can be directly written for a chosen small value of $\Delta t$ to obtain $\mu_0$:

$$\mu_0(x_i) = \hat{\rho}_{\text{eq}}(x_i)e^{\Delta t \hat{R}_i} = \hat{\rho}_{\text{eq}}(x_i)(1 + \Delta t \hat{R}_i + \mathcal{O}(\Delta t^2)). \tag{10}$$

Next, the effect of the transport step (5) is to rearrange mass via diffusion and drift so that we return to the steady state distribution $\hat{\rho}_{\text{eq}}$. We cannot take $\mu_1$ to be $\hat{\rho}_{\text{eq}}$ exactly, since a single step of the splitting scheme introduced in (4) and (5) is only accurate up to $\mathcal{O}(\Delta t^2)$. Therefore, a straightforward application of the growth step (10) will result in a slight change in the total mass of the system. Additionally, in practice we have only estimates $\hat{R}_i$ of the true flux rates, further contributing to this effect. Consequently, we must instead re-normalise $\mu_1$ so that it has the same mass as $\mu_0$:

$$\mu_1(x_i) \propto \hat{\rho}_{\text{eq}}. \tag{11}$$

With $\mu_0$ and $\mu_1$ constructed in this way, we may compute the solution to the discrete Schrödinger problem (8).

**Choice of $\varepsilon$ and $\Delta t$** The key parameters for the scheme we describe are $\Delta t$, the time step introduced in the growth splitting, and the regularisation parameter $\varepsilon$ for entropy-regularised optimal transport. In the theoretical framework of the Schrödinger problem, these parameters have a proportional relationship $\varepsilon = \sigma^2\Delta t$. The accuracy of the scheme should improve in the limit as $\varepsilon \to 0, \Delta t \to 0$ and $\varepsilon = \sigma^2\Delta t$, since the splitting approximation becomes exact. However, in practice where we have discrete samples we find that allowing $\varepsilon$ and $\Delta t$ to deviate from this relationship often leads to better results.

In the discrete setting, a key limitation is the value of $\varepsilon$, which controls the level of diffusion in the reference process in the Schrödinger problem, and consequently influences the level of diffusion in the inferred process. In practice when we are dealing with a limited number of samples in a potentially high-dimensional space, taking $\varepsilon$ too small may lead to an ill-conditioned problem. The reason for this is that the distance between points in the set of samples $\overline{\mathcal{X}}$ may be quite large compared to $\varepsilon = \sigma^2\Delta t$. That is, $\exp(\frac{-\|x-y\|^2}{2\sigma^2\Delta t})$ may be exceedingly small, resulting in a reference process that mixes extremely slowly. On the other hand, if we pick a reasonably sized $\varepsilon$, strictly adhering to the proportionality relationship may mean that the corresponding $\Delta t$ is too large for the splitting approximation to be a good one. In practice, we have often found that it is helpful to take $\Delta t$ to be slightly smaller (and consequently $\varepsilon$ slightly larger) than what is expected in theory. We discuss this choice in practice further in Section 2.6.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## 2.4 Quadratically regularised optimal transport

The entropy-regularised optimal transport problem (9) is well known for its probabilistic interpretation and the existence of an efficient solution scheme by matrix scaling [25]. However, the use of entropic regularisation results in a transport plan that necessarily has a dense support [2]. Recent contributions to the optimal transport literature [22, 2] have highlighted that alternative choices of the regulariser may yield other smooth approximations of the Monge-Kantorovich problem which exhibit desirable properties. In particular, using a quadratic ($L^2$) regulariser to form the problem

$$\min_{\gamma \in \Pi(\mu_0, \mu_1)} \int C(x, y) d\gamma(x, y) + \varepsilon \|\gamma\|_2^2 \tag{12}$$

gives rise to what is known as the *quadratically* regularised optimal transport problem. As noted by [22, 2], quadratically regularised OT has the property that transport plans are generally sparse in practice (in the discrete case, transition probabilities are nonzero only on a sparse graph that spans the data), making it a favourable choice for interpretability of transport plans as well as computational efficiency (see Section 2.9). In addition, in [2] the authors remark that the quadratically regularised problem may be less prone to issues of numerical stability.

In practice, we may employ a quadratic regularisation in our scheme by substituting the solution of (12) for the optimal coupling $\gamma_{\Delta t}$ instead of the entropy-regularised problem (8). As we demonstrate in Sections 2.6 and 2.7, we find evidence that quadratic regularisation is more robust to parameter choices and noise compared to entropy regularisation.

## 2.5 Extension to non-potential vector fields

Estimation of dynamics in the case where the underlying drift $\mathbf{v}$ does not arise from a potential gradient requires additional information to be available, such as potentially noisy or partial estimates of the velocity of cells [16, 1, 17]. Since at its core our method is based on solving a convex optimisation problem, additional information such as velocity estimates can be incorporated into our estimation procedure in a straightforward manner by modifying the cost matrix $C$. Indeed, suppose for each observed cell $x_i$ we also have an estimate of its velocity $\mathbf{v}_i$. In the setting of velocity estimates derived from RNA velocity, the orientation of velocity estimates is more biologically informative than the magnitude [35], and it is therefore natural to incorporate velocity information in terms of cosine similarities [1, 17]. In our case, we consider an overall cost function that is a linear combination of the standard squared Euclidean cost $C_{\text{euc}}$ and a matrix of cosine similarities $C_{\text{velo}}$, i.e.

$$C = \lambda_1 C_{\text{euc}} + \lambda_2 C_{\text{velo}},$$

where

$$(C_{\text{velo}})_{ij} = \frac{1}{2}\left(1 - \frac{\langle \mathbf{x}_j - \mathbf{x}_i, \mathbf{v}_i \rangle}{\|\mathbf{x}_j - \mathbf{x}_i\| \|\mathbf{v}_i\|}\right). \tag{13}$$

In practice, the weights $\lambda_1, \lambda_2$ would depend on the relative scales of $C_{\text{euc}}$ and $C_{\text{velo}}$, as well as any cost normalisation that is applied. We discuss the details of this in a simulated example in Section 2.7.

## 2.6 Simulated data – potential driven dynamics

**Simulation setup and parameters** We first consider a tri-stable system (1) in $\mathcal{X} = \mathbb{R}^{10}$, with drift term $\mathbf{v}$ taken to be the negative gradient of the potential

$$\Psi(x) = 2.5\|x - z_0\|^2 \|x - z_1\|^2 \|x - z_2\|^2, \tag{14}$$

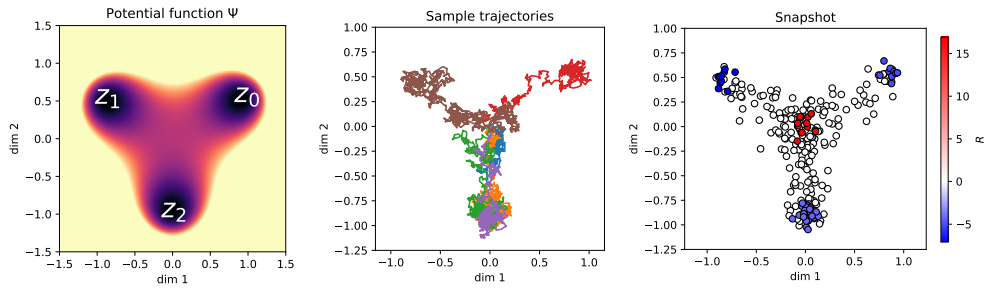
with wells $\{z_0, z_1, z_2\}$ located at

$$\begin{aligned} z_0 &= 1.05\,[\cos(\pi/6), \sin(\pi/6), 0, \ldots, 0]^\top \\ z_1 &= 1.05\,[\cos(5\pi/6), \sin(5\pi/6), 0, \ldots, 0]^\top, \\ z_1 &= 1.05\,[\cos(\pi/2), \sin(\pi/2), 0, \ldots, 0]^\top. \end{aligned}$$

We illustrate this potential landscape in Figure 3(a) in the first two dimensions of $\mathcal{X}$. Simulated particles are initially isotropically distributed around the origin following the law $X_0 \sim 0.01\mathcal{N}(0, I)$ at $t = 0$, where

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\mathcal{N}(0, I)$ denotes the standard normal distribution in $\mathbb{R}^{10}$ with covariance $I$. Particles then evolve following drift-diffusion dynamics with $\sigma^2 = 0.5$. Whenever a particle falls in the vicinity of any of the potential wells $\{z_0, z_1, z_2\}$, it is removed with exponential rate 5. That is, in each time step $\mathrm{d}t$, a particle located in a sink region is removed with probability $5\,\mathrm{d}t$. We defined the sink region for each potential well $z_i$ to be a ball of radius $r = 0.25$ centred at $z_i$.




(a) (b) (c)

Figure 3: (a) Illustration of the potential $\Psi$ in the first two dimensions of the space $\mathcal{X}$. (b) Examples of simulated particle trajectories $X_t^{(i)}$ following the drift-diffusion process. (c) Snapshot particles $\hat{\rho}_{\mathrm{eq}}$ shown in the first two dimensions of $\mathcal{X}$, with the value of $R$ indicated. Source and sink regions correspond to $R > 0$ and $R < 0$ respectively.

Exact sampling of snapshots from the steady state distribution $\rho_{\mathrm{eq}}$ of (2) would require the solution of a high-dimensional PDE and is therefore computationally difficult. Instead, we obtain an approximate snapshot of the system at its steady state by simulating $N = 250$ trajectories from start to finish using the Euler-Maruyama method

$$X_{t+\tau} = X_t - \tau \nabla \Psi(X_t) + \sigma\sqrt{\tau}\mathcal{N}(0, 1). \tag{15}$$

For our simulations we employed a time step $\tau = 1 \times 10^{-3}$, and from each trajectory $\{X_t^{(i)} : 0 \leq t \leq T_{\mathrm{final}}^{(i)}\}, 1 \leq i \leq N$ we sampled a single particle state chosen at a random time chosen uniformly on $[0, T_{\mathrm{final}}^{(i)}]$ to form the snapshot data $\hat{\rho}_{\mathrm{eq}}$. This scheme was also the one used for obtaining snapshots in [41].

Particles $x_i$ located in the sink regions were labelled as 'sink' sites and assigned flux rates $R_i$ so that the average sink flux rate was $-5$ and total flux rate for each well $\{z_0, z_1, z_2\}$ matched the ground truth in proportion. Particles located in a ball of radius 0.25 of the origin were labelled as 'source' sites, corresponding to locations $x_i$ with $R_i > 0$. Since we deal with finite samples, we assigned $R_i$ uniformly on source sites such that the equilibrium condition $\sum_i R_i = 0$ was satisfied.

We display some example trajectories in Figure 3(b), and illustrate the snapshot data $\hat{\rho}_{\mathrm{eq}}$ in Figure 3(c), where the values of $R_i$ at source and sink sites are shown by colour.

**Inferring dynamics using StationaryOT** To apply StationaryOT, we chose a time step $\Delta t = 25\tau = 2.5 \times 10^{-2}$, noting that this is small compared to the average particle lifespan of 0.934 in this simulation. We solved the StationaryOT problem using entropy-regularised optimal transport as described in Section 2.3, using a range of regularisation parameter values $\varepsilon$ in $10^{-2.5} - 10^{1}$. As we discuss in more detail later, we found that $\varepsilon = 0.026$ best matched the ground truth in terms of average fate probability correlation across the three lineages. For this choice of $\varepsilon$ we computed a forward transition matrix $P$ from the optimal transport coupling $\gamma_{\Delta t}$ by row-normalising:

$$P_{ij} = \frac{(\gamma_{\Delta t})_{ij}}{\sum_j (\gamma_{\Delta t})_{ij}}.$$

The matrix $P$ therefore describes a time-$\Delta t$ evolution of probability densities on the discrete set $\overline{\mathcal{X}}$. For an initial distribution $\pi_0$ supported on $\overline{\mathcal{X}}$, we can compute the evolution $\{\pi_0 P^k, k = 0, 1, 2, \ldots\}$ over steps of length $\Delta t$, which we take to be an estimate of the dynamics of the underlying drift-diffusion process. In Figure 4 we show the inferred process for $k = 1, 5, 10, 20$ where we have taken $\pi_0$ to be uniform on the source sites.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

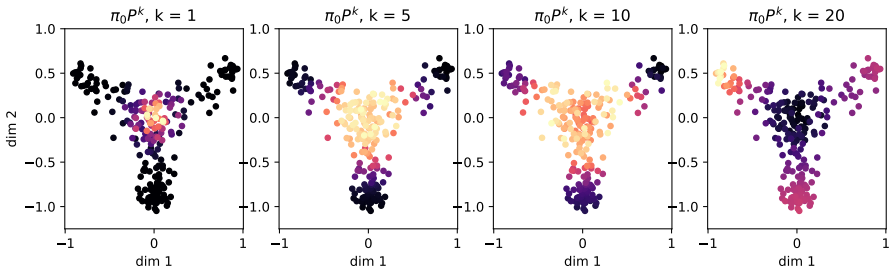



Figure 4: Evolution of the dynamics recovered by StationaryOT.

From the transition probabilities $P_{ij}$ we may compute fate probabilities for each of the three lineages defined by the potential wells $\{z_0, z_1, z_2\}$. (These are absorption probabilities of the Markov chain $P$ – see Section A for details). We summarise these fate probabilities in Figure 5, and find that the correspondence between inferred and ground truth fate probabilities measured in terms of the Pearson correlation is high ($r \approx 0.99$). As another measure of the accuracy of the estimated dynamics, we compute the mean-first passage time (MFPT) of each sampled point $x_j$. This is the expected time at which a Markov chain initialised at a randomly chosen source location $x_i$ reaches $x_j$:

$$\mathrm{MFPT}(x_j) = \mathbb{E}_{x_i \sim \mathrm{sources}} \, \mathrm{MFPT}(x_j | x_i),$$

where $\mathrm{MFPT}(x_j|x_i)$ denotes the conditional MFPT for a particle starting at $x_i$ to hit state $x_j$. Comparing the MFPT estimates to the ground truth MFPT in Figure 6(a), we find that the correspondence is high ($r > 0.9$).

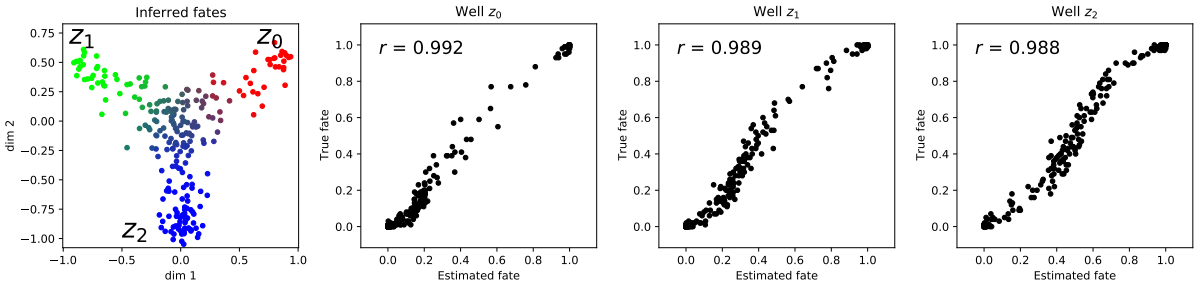



Figure 5: Visualisation of estimated fate probabilities towards each of the wells $\{z_0, z_1, z_2\}$ on the snapshot coordinates, as well as correlation with ground truth fate probabilities.

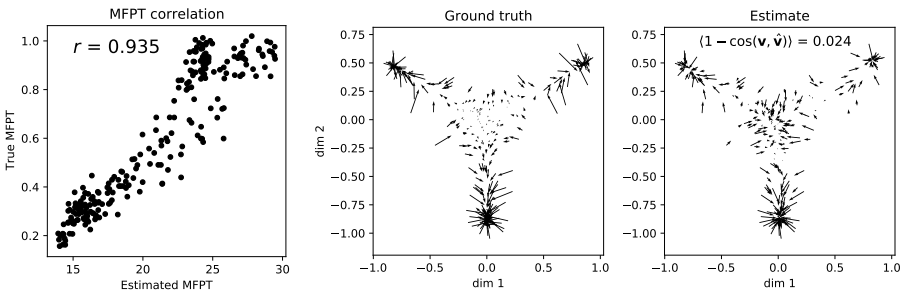



Figure 6: (a) Comparison of estimated MFPT to ground-truth MFPT. (b) Comparison of recovered velocities to ground truth velocity.

**Reconstructing the drift field v** Since the transition probabilities encode the displacement law of the underlying process over a time interval $\Delta t$, we can also recover an estimate $\hat{\mathbf{v}}$ of the velocity field $\mathbf{v}$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by computing the expected time-$\Delta t$ displacement of each cell:

$$\hat{\mathbf{v}}(x_i) = \frac{\mathbb{E}_P(X_{\Delta t} - X_0 | X_0 = x_i)}{\Delta t}.$$

In Figure 6(b) we show the estimated velocity field $\hat{\mathbf{v}}$ alongside the ground truth $\mathbf{v}$, and we measure the error by computing the mean cosine error between vector fields:

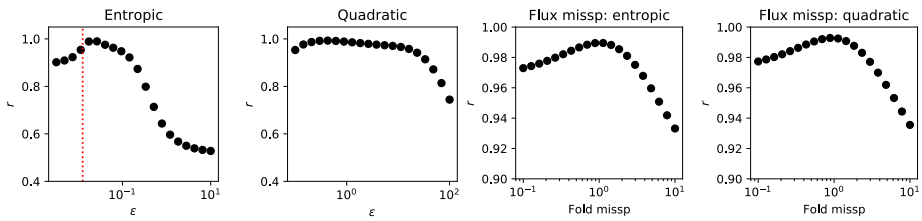
$$\frac{1}{N}\sum_{i=1}^{N}(1 - \cos\angle(\mathbf{v}(x_i), \hat{\mathbf{v}}(x_i))) \approx 0.024.$$

We observe that the estimated field $\hat{\mathbf{v}}$ resembles the ground truth quite well near the potential wells where particles are subjected to a relatively strong drift, but struggles near the origin where the true velocity field has a small magnitude. Overall however, the cosine error is close to zero, indicating that our recovered velocity field matches the ground truth field well.

**Effect of the choice of regularisation parameter $\varepsilon$ and flux rates $R$** We next turn to investigating the effects of the choice of the regularisation parameter $\varepsilon$ on the quality of the recovered dynamics. To quantitatively measure this, we choose to compute the average correlation $r$ between estimated and ground truth fate probabilities across the three lineages. We applied StationaryOT using both entropy and quadratic regularisations, and let $\varepsilon$ vary on a log-scale from $10^{-2.5} - 10^{1}$ and $10^{-1} - 10^{2}$ respectively.

As shown in Figure 7, in the case of entropy-regularised optimal transport we observe in both the fate and velocity estimates that there is clearly a single optimal value of this parameter at $\varepsilon = 0.026$. This is larger than the theoretically optimal value of $\sigma^2\Delta t = 0.0125$, in keeping with our observations discussed in Section 2.3. However, as we can see in Figure 7(a), StationaryOT with the theoretically optimal value fares only slightly worse and is located close to the maximum. When $\varepsilon$ is chosen too small or too large, performance degrades. On the other hand, we find that performance when using a quadratic regularisation is much less sensitive to the choice of $\varepsilon$, with the correlation over $\varepsilon$ showing a much flatter profile.

Since flux rates $R$ are also parameters that need to be specified, we examine the sensitivity to varying the flux rate in Figure 7. Here, we over- and underestimate the proportion of particles that leave the system through the well $z_0$ by up to 10 fold, and we observe that performance degrades moderately in either case. We show results for entropic and quadratic regularisation where $\varepsilon$ is chosen to be the optimal values of 0.026 and 0.43 respectively, and note that both choices of regularisation behave virtually identically.




(a) (b)

Figure 7: Summarised performance for (a) varying regularisation parameter $\varepsilon$ for entropy and quadratic regularisations and (b) misspecification of flux rates. In (a), the theoretically optimal value of $\varepsilon$ is indicated in red.

**Laws on paths** The stochastic differential equation (1) naturally induces a probability measure on the space of continuous functions valued in $\mathcal{X}$, from which one can sample cell trajectories. We discuss this point of view at length in related work on the non-stationary case [18]. From this perspective, we may treat the recovered process as inducing a law on discrete-time paths valued in $\overline{\mathcal{X}}$, and we expect that a good estimate of the dynamics should correspond to a law on paths that is closer to the ground truth law. To illustrate this, in Figure 8 we display 100 sample paths over 25 timesteps, i.e. $t \in \{0, \Delta t, 2\Delta t, \ldots, 24\Delta t\}$. We compare the ground truth to the StationaryOT output for both entropic and quadratic OT for optimal and sub-optimal (taken as 10× the optimal value) choices of the regularisation parameter $\varepsilon$. Visually, it is clear that StationaryOT using both entropic and quadratic

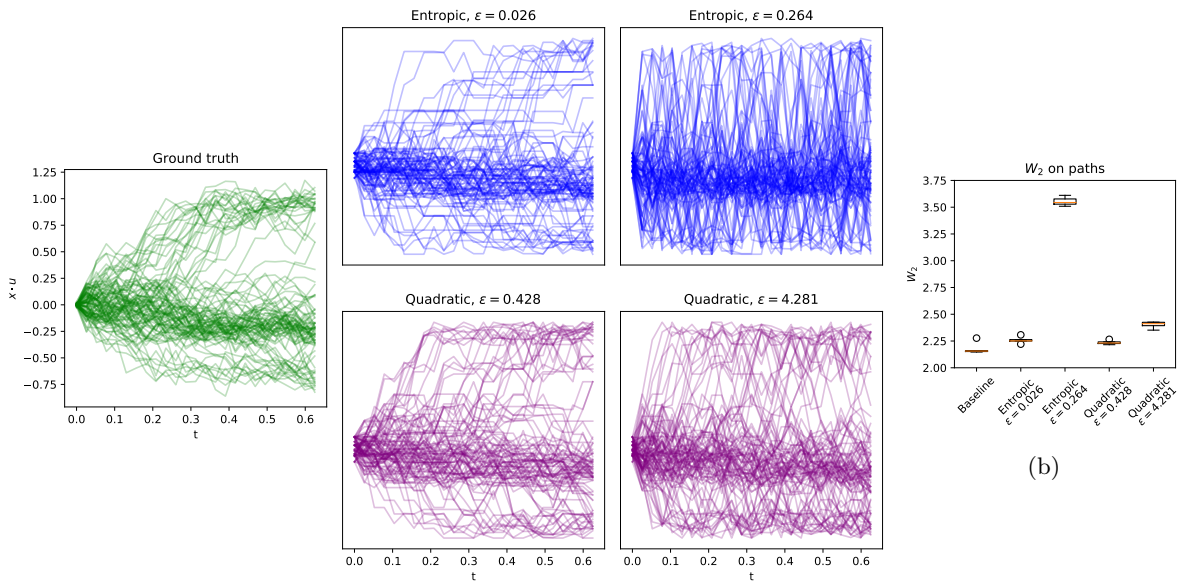
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OT produces very similar output resembling the ground truth when $\varepsilon$ is chosen to be optimal. On the other hand, when $\varepsilon$ is chosen to be too large we observe a visible worsening of performance, with more paths jumping between branches. As we also observed in terms of fate probabilities, the performance of StationaryOT with quadratic regularisation appears to degrade more gracefully than entropic OT.

To provide a quantitative assessment of performance, the natural metric to use is the 2-Wasserstein ($W_2$) distance on the space of laws on paths, as we also argue in [18]. Since we work in the setting of discrete time and space, we take the ground metric on $\overline{\mathcal{X}}^T$ to be the squared $L^2$ distance:

$$\|f - g\|_2^2 = \sum_{i=1}^{T} \|f_i - g_i\|_2^2.$$

Using the $W_2$ metric for laws on paths, we computed the error of each reconstruction relative to the ground truth and display the results in Figure 8(b). Importantly, we note that since we are dealing with finite samples, the expected $W_2$ distance between independent collections of sample paths from the same distribution will be nonzero. Thus, as done in [18] we compute a baseline error as the $W_2$ distance between independent samplings of 250 paths from the ground truth. In Figure 8(b) we show the average $W_2$ error over 5 resamplings of 250 paths, from which we note that StationaryOT with entropic or quadratic OT yields results that are close to the baseline in $W_2$ error when $\varepsilon$ was chosen to be optimal. On the other hand, picking $\varepsilon$ to be too large leads to a higher error for both methods, but with entropic OT performing significantly worse than quadratic OT.




(a) (b)

Figure 8: (a) Collections of 100 sample paths from the ground truth process (1) as well as StationaryOT outputs for both entropic and quadratic OT with optimal and sub-optimal $\varepsilon$. (b) $W_2$ error on paths for StationaryOT reconstructions, shown for 5 repeated samplings of 250 paths.

## 2.7 Simulated data – non-conservative dynamics

Now we consider the case where the drift $\mathbf{v}(x)$ is no longer the gradient of a potential landscape, i.e. there is a curl component. In this case, the underlying process is no longer identifiable from only sampled spatial locations [41, 18], and it is necessary to have additional velocity estimates in order to estimate cellular trajectories.

**Simulation setup and parameters** To illustrate this, we consider a process with a drift field given by the sum of a potential-driven term and a non-conservative vector field, i.e.

$$\mathbf{v}(x) = -\nabla\Psi(x) + \mathbf{f}(x). \tag{16}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Again, we work in $\mathcal{X} = \mathbb{R}^{10}$ and we take

$$\Psi(x) = \exp\left(-\frac{x_1^2 + x_2^2}{h^2}\right) + \frac{1}{2}(x_1^2 + x_2^2) + 10\sum_{i=3}^{10} x_i^2 \tag{17}$$

$$\mathbf{f}(x_1, x_2) = 10\exp\left(-\frac{x_1^2 + x_2^2}{h^2}\right)\begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}. \tag{18}$$

We pick $h = 0.5$, controlling how rapidly the field $\mathbf{f}$ decays and the location of the potential well in $\Psi$. In the first two dimensions of $\mathcal{X}$, particles can be thought of as diffusing on a radially symmetric potential field with a ring of wells located about the origin, and subject to a superimposed anticlockwise vector field that decays away from the origin. We show a surface plot of $\Psi(x)$ and a vector field plot of $\mathbf{f}(x)$ in Figure 9(a).

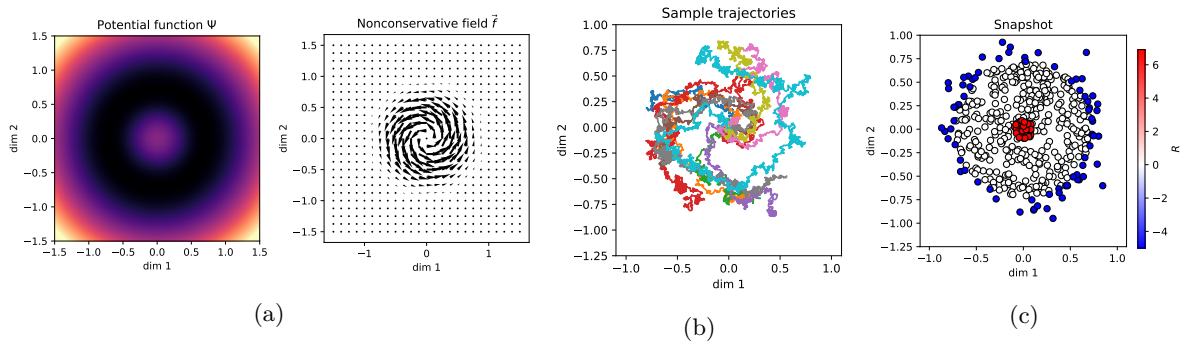



Figure 9: (a) Illustration of potential-driven ($\Psi$) and non-conservative ($\mathbf{f}$) components of the overall drift $\mathbf{v}$. (b) Examples of simulated particle trajectories $X_t^{(i)}$ following the drift-diffusion process. (c) Snapshot samples shown in the first two dimensions of $\mathcal{X}$, with source ($R > 0$) and sink ($R < 0$) regions indicated.

We initialise particles following the initial distribution $X_0 \sim 0.01\mathcal{N}(0, 1)$ that are then subject to the drift-diffusion process with diffusivity $\sigma^2 = 0.1$. The minimum of the circular potential well is located along a cylinder of radius 0.721 about the origin in the first two dimensions of $\mathcal{X}$, and we treat all points outside this cylinder as a sink region, in which particles are removed at exponential rate 5. We sample 500 particles from this process and designate cells found within a ball of radius 0.1 about the origin to be source cells, and cells located in the sink region to be sink cells. Sink cells were assigned a flux rate $R_i = -5$, and source cells were assigned uniformly so that $\sum_i R_i = 0$ as we did previously in Section 2.6. We illustrate in Figure 9(b) some example trajectories from this simulation, and in Figure 9(c) we display the sampled snapshot $\hat{\rho}_{\text{eq}}$ along with the flux rates.

**StationaryOT with and without velocity data** For each sampled cell $x_i$, we obtain velocity estimates by evaluating the drift vector field $\mathbf{v}(x_i)$ at its location. We then formed two cost matrices: $C_{\text{euc}}$, the matrix of squared Euclidean distances, and $C_{\text{velo}}$ the matrix of cosine similarities as defined in (13). Both matrices were normalised to have unit mean. Note here that this normalisation is purely an empirical choice, and no corresponding normalisation of the cost was performed in Section 2.6 because of the theoretical motivation in the potential-driven case.

We constructed the optimal transport cost matrix to be a convex combination of the Euclidean and velocity cost matrices:

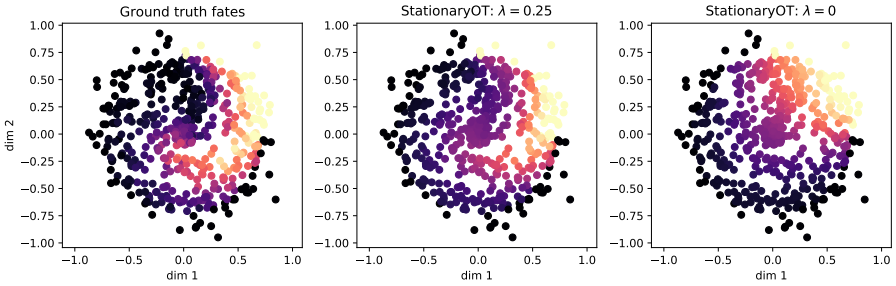
$$C = (1 - \lambda)C_{\text{euc}} + \lambda C_{\text{velo}},$$

and we took $\lambda = 0.25, 0$, respectively corresponding to StationaryOT with and without velocity information. Both entropic and quadratic OT were used to solve for couplings, with $\varepsilon = 0.05$ and $\varepsilon = 0.5$ respectively. Since the setting of this simulation is rotationally invariant in the first two dimensions, we choose to summarise our results in terms of the absorption probabilities for cells entering the region

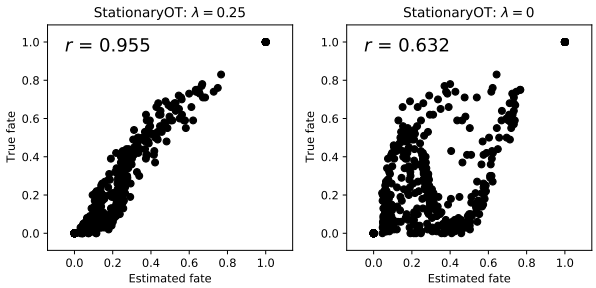
$$\mathcal{S} = \{x_i : \Theta(x_i) \in [0, \pi/2] \text{ and } R_i < 0\},$$

i.e. the set of sink cells in the first quadrant in $(x_1, x_2)$. As shown in Figure 10(a), the ground truth fate probabilities clearly capture the rotational component of drift, with the set of cells fated towards $\mathcal{S}$ forming a curled shape. We observe that StationaryOT with velocity data produces results qualitatively

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




(a)




(b)

Figure 10: (a) Comparison of fate probabilities towards the sinks in the first quadrant. (b). Correlation of estimated fate probabilities to ground truth fates with and without incorporation of velocity data.

capturing this effect, whilst neglecting velocity information leads to a symmetric fate profile that reflects only the potential-driven component as expected. To quantitatively compare fates, we computed as previously the Pearson correlation between the estimated fate probabilities and the ground truth. We show this in Figure 10(b), from which we observe that StationaryOT with velocity data produces a markedly improved fate correlation ($r = 0.953$) compared to StationaryOT without velocity data ($r = 0.631$).

**Laws on paths** As we have done in Section 2.6, we may examine sample paths from the ground truth process as well as the estimates output by StationaryOT. We sample trajectories with the initial condition

$$\pi_0 = \{x_i : \Theta(x_i) \in (-\pi/6, \pi/6) \text{ and } \|x_i\| \in (0.25, 0.5)\}.$$

We illustrate these in Figure 11 in the first two dimensions of $\mathcal{X}$. Again, we observe that incorporation of velocity estimates yields results that clearly reflect the rotational trajectories in the ground truth. On the other hand, without using velocity information, we observe sample paths consistent with only the potential-driven component. Additionally, for either choice of regularisation we observe that StationaryOT overestimates the rotational drift as cells settle into the potential well. This effect can be attributed to the fact that the cosine similarity cost (13) depends only on the orientation of the rotational field, and thus is unaware of its decay as cells drift towards the well. In this situation, we can only expect to capture the rotational field qualitatively rather than quantitatively. We suggest that possible remedies for this effect may include weighting entries of $C_{\text{velo}}$ by velocity magnitudes or using an alternative velocity cost that is based on squared Euclidean distances.

**Sensitivity to noise** Finally, are interested in investigating the behaviour of StationaryOT when the provided velocity estimates are subject to additive noise, that is

$$\hat{\mathbf{v}}(x_i) = \mathbf{v}_0(x_i) + C\eta\mathcal{N}(0, I)$$

where $C$ is a scale constant chosen such that $\mathbf{v}_0/C$ has order 1, i.e. the noise term is on the same order as the signal. We pick $C = \mathbb{E}_{x_i}\|\mathbf{v}_0(x_i)\|$. We applied StationaryOT using both entropic and quadratic OT for values of $\eta \in [0, 2]$ and choices of regularisation $\varepsilon$ chosen in the range $10^{-2} - 10^0$ (logarithmic) for entropic OT and $0.5 - 10$ (linear) for quadratic OT.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

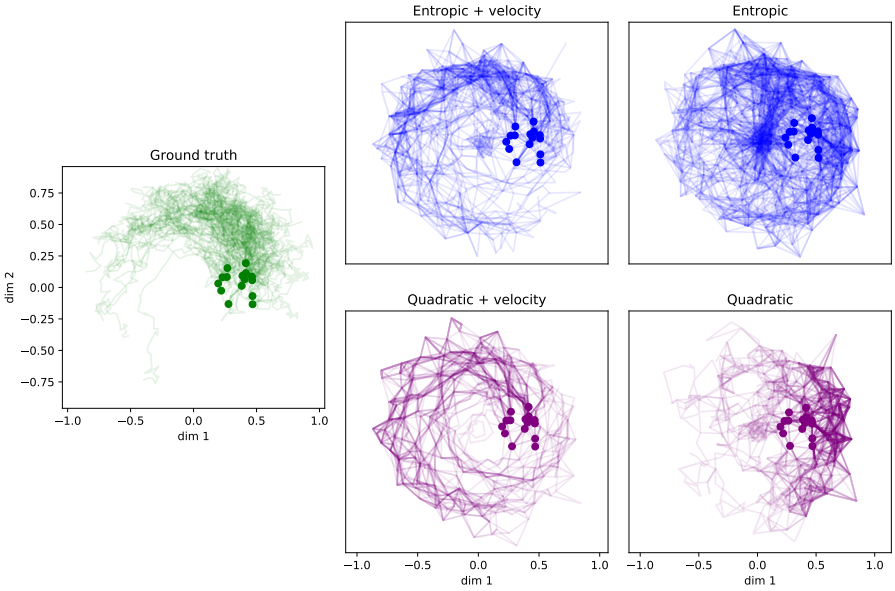



Figure 11: Collections of 100 sample paths drawn from the ground truth process (1), as well as StationaryOT output with and without velocity estimates for both entropic and quadratic OT. We indicate the initial condition $\pi_0$ as dots.

For additional comparison, for each noise level $\eta$ we also computed a transition matrix based solely on cosine similarities of velocity estimates to $k$-nearest neighbour ($k$-NN) graph edges using the scVelo package [1] in which the transition law for each cell $x_i$ is

$$\mathbb{P}[X_{\Delta t} = x_j | X_0 = x_i] \propto \exp\left(k \cos \angle(x_j - x_i, v_i)\right), \quad x_j \in \text{neighbours}(x_i). \tag{19}$$

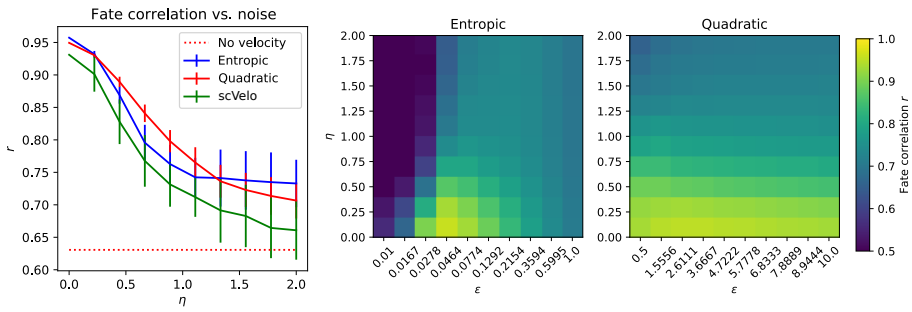
In the above, $k$ is a scale parameter controlling the level of directedness in the resulting transition law, with larger $k$ corresponding to increased directedness in the transition law. We used $k$ in the range $2.5 - 25$ and all other parameters were taken to be defaults.

In each case, performance was summarised as we did previously in terms of the fate correlation for the set $\mathcal{S}$. We show results summarised over 10 independent repeats in Figure 12 and we observe that, as expected, performance degrades for all methods as the level of noise increases. However, StationaryOT with either entropic or quadratic regularisation consistently produces more accurate fate estimates compared to the scVelo method. We argue that this effect reflects the fact that StationaryOT is a global method and solves for transition laws that best agree with the inputs across the dataset. On the other hand, the scheme (19) is local in that the transition law for a single cell can be determined by only considering a single velocity vector and a few neighbouring locations.

Additionally, for moderate levels of noise ($\eta \in [0, 1]$), we observe that the quadratic regularisation outperforms the entropic regularisation. Roughly speaking, we suspect that this results from the fact that the transition laws recovered by quadratic OT are concentrated on a sparse subset of cells, limiting the effect that an error in any single velocity estimate can have on the overall Markov chain. On the other hand, since entropic regularisation yields dense transition laws, errors can be propagated across the full support of $\overline{\mathcal{X}}$.

In Figure 12(b) we examine the sensitivity of StationaryOT performance on the choice of the regularisation $\varepsilon$. As we observed in the case of Section 2.6, the entropic regularisation appears to depend strongly on the choice of $\varepsilon$ whereas a quadratic regularisation behaves similarly across the values of $\varepsilon$ used.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

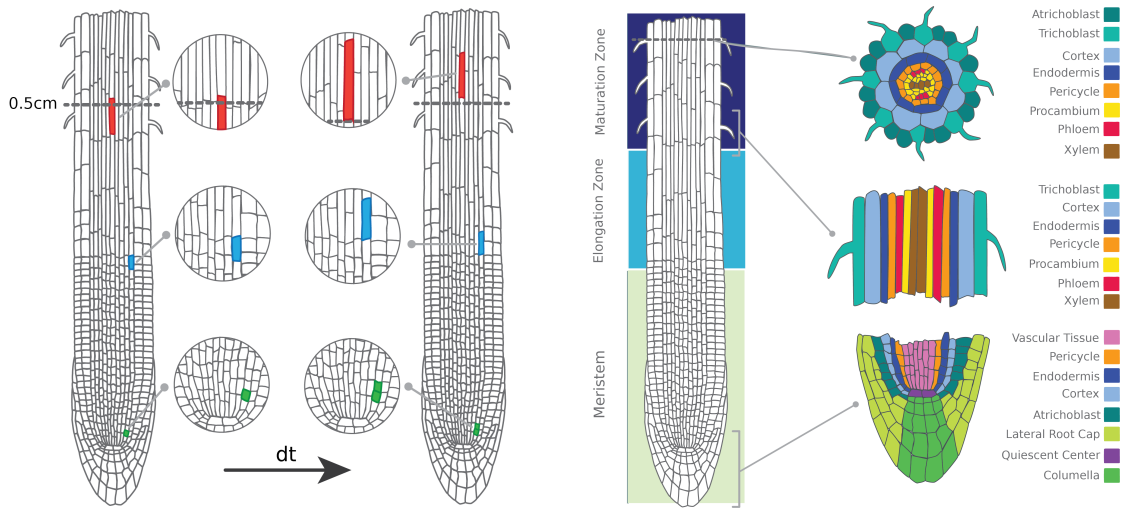



(a) (b)

Figure 12: (a) Correlation of estimated fate probabilities to ground truth as a function of noise $\eta$. (b) Sensitivity of entropic and quadratic regularisations to the choice of $\varepsilon$.

## 2.8 *Arabidopsis thaliana* root tip scRNA-seq data

**Overview** We now apply StationaryOT to the scRNA-seq atlas dataset generated by Shahan et al. [30], which comprises of gene expression data from $1.1 \times 10^5$ cells from the first 0.5 cm of the *Arabidopsis thaliana* root tip. Stem cells occur close to the tip of the root and differentiate into ten distinct lineages (see Figure 13), with cells becoming increasingly differentiated as they increase in distance from the stem cells. Additionally, the terminal 0.5 cm of the root captures all tissue developmental zones, including the root cap, meristem, elongation zone, and part of the maturation zone. While new cells are constantly produced in the meristem, the bottom 0.5 cm is expected to be in equilibrium as cell division and elongation push existing differentiated cells out of the 0.5 cm section of interest, preserving a constant profile of cell types as illustrated in Figure 13a. We also show the root tip anatomy diagrammatically in Figure 13b, indicating developmental zones and tissue types.




(a) (b)

Figure 13: (a) While individual cells divide (green), elongate (blue), and are displaced from the bottom 0.5 cm (red) as the root grows, cell populations remain in equilibrium. (b) The structure of the *Arabidopsis thaliana* root tip by developmental zone (left) and lineage (right) (Illustrations modified from the Plant Illustration repository [26]).

**Application of StationaryOT** For each cell $x_i$, daily growth rates $g_i$ were estimated from imaging data of the growing meristem over a week-long period [28]. Using these growth rates and the proportion of cells expected to be actively dividing, we estimated that roughly 5% of the cells in each lineage would be replaced in a 6-hour period ($\Delta t = 0.25$) and selected the 5% of most differentiated cells from each lineage as sinks, as defined by pseudotime. For these sink cells we set $g_i = 0$, i.e. they are completely removed over the time interval. The remaining cells were set as sources, with $R_i$ chosen to agree with

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the biological growth estimates, i.e. for each cell $x_i$ we take $R_i$ such that $\exp(R_i) = g_i$.

We applied StationaryOT using both entropic and quadratic regularisations with parameters $\varepsilon = 0.025\overline{C}$ and $\varepsilon = 2.5\overline{C}$ respectively, where the scale factor $\overline{C}$ is taken as the mean value of the squared Euclidean cost matrix $C$. We found our results to be robust to changes of a factor of two in the number of sinks, the time step size, and $\varepsilon$. Due to computational limitations of the standard implementation of the method, we applied StationaryOT to batches of 10,000 cells sampled from the full dataset, though in Section 2.9 we demonstrate two methods to scale the analysis to the full atlas.

In *Arabidopsis* root development, cell lineage is fixed early in development [28]. Thus, for each cell $x_i$ we may regard the lineage $j$ corresponding to the largest fate probability, i.e. $\operatorname{argmax}_j p_{ij}$ as its putative fate. We checked whether these putative fates matched the manually curated atlas annotation, and used the magnitude of the corresponding fate probability, $0 \leq p_{ij} \leq 1$, as a measure of the confidence of prediction. StationaryOT with quadratic and entropic regularisation performed similarly in terms of the percentage of cells where the putative fate matched the atlas annotation, matching 81% and 80% of cells respectively (see Figure 15). Both regularisations also performed similarly in terms of the magnitudes of the putative fates, with the entropic regularisation achieving an average of 69%, increasing from an average of 40% for cells in the meristem to an average of 84% for cells in the maturation zone and quadratic achieving an average of 65%, ranging from 38% in the meristem to 80% in the maturation zone (see Figure 23).

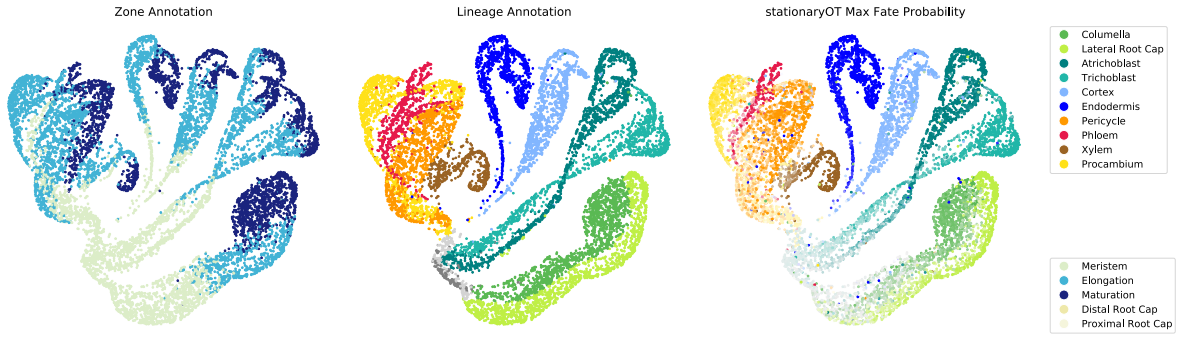



Figure 14: Developmental zone (left) and lineage annotations (centre) shown on a UMAP embedding. Putative fate probabilities from StationaryOT with entropic regularisation are visualised on the right, where each cell is coloured by putative fate and its saturation based on the magnitude of that probability. For over 80% of cells the putative fate matched the annotation, with the magnitude of the probability increasing later in development.

Both choices of regularisation performed well on nine of the ten lineages, struggling only with mature procambium cells, as shown in Figure 15. We believe this is due to an inconsistency between the pseudotime and developmental zone annotations, where cells in the elongation zone received higher pseudotimes than cells in the maturation zone, resulting in them being incorrectly set as terminal states (See Figure 20). Both regularisations were robust to changes in parameters, where the percentage of cells whose putative fate matched the annotation changed by no more than 2% when changing a parameter by a factor of two. StationaryOT with quadratic regularisation was particularly robust, with performance degrading by no more than 2% when multiple parameters were changed by up to a factor of five (See Section B.3).

**Comparison to PBA** Since population balance analysis (PBA) [41] addresses the same problem as StationaryOT (see Section 1.4), it is natural to evaluate its performance on the *Arabidopsis* root dataset. We show results for 1% sinks, $\Delta t = 0.25$ (6-hour time step), diffusivity $D = 2.5$, and $k = 10$ for the $k$-NN graph. These parameters were found to yield the best results over a parameter sweep (see Section B.3). PBA was on par with the StationaryOT methods, with 81% of putative fates matching the annotation, compared to 80% and 81% for the StationaryOT analyses (See Figure 15). Average fate probabilities were also similar, with PBA achieving an average of 64% compared to 65% and 69% for the StationaryOT methods (see Figure 23). Given that PBA and StationaryOT are methodologically distinct, the fact that they perform similarly is a strong indication that the results reflect the underlying biology, rather than artefacts from the respective models.

In general however, we found PBA to be more sensitive to parameter values than StationaryOT.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

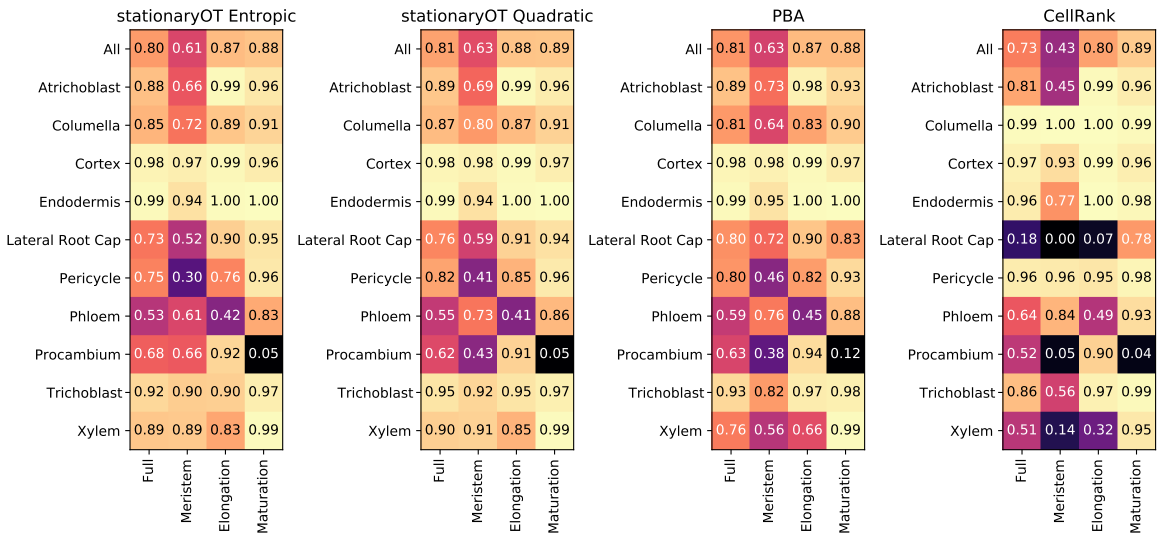



Figure 15: Proportion of cells where the maximum probability matches the annotation by developmental zone and lineage.

Assigning 5% of cells in each lineage as sinks for a 6-hour time step ($\Delta t = 0.25$) is biologically motivated in order to balance the number of cells created due to growth with the number of sinks. Using this sink selection scheme, PBA was found to perform poorly for the columella lineage, incorrectly assigning all columella cells a putative fate of lateral root cap. This lowered the overall percentage of putative fates that matched the annotation to 72% (See Figure 19a). Only through an extensive parameter sweep did we find the combination of parameters that resulted in columella cells receiving the correct putative fate. We found PBA to be generally more sensitive to other parameter changes, matching 7% fewer cells to the annotation when a single parameter was changed by a factor of two compared to only 2% for the StationaryOT methods (See Section B.3). This may be of concern, since in many applications there may not be sufficient prior biological knowledge to distinguish between good and bad parameter choices.

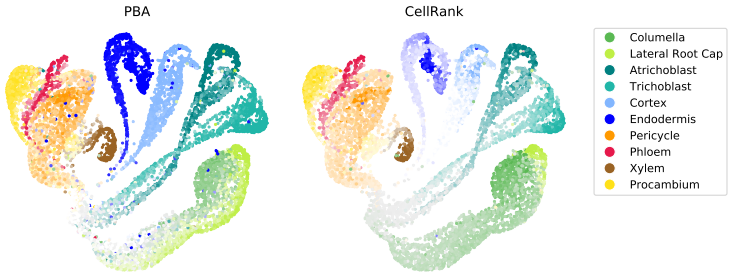
**Comparison to CellRank** CellRank is a trajectory inference method that uses both transcriptomic similarity and RNA velocity data to estimate transition laws for cells [17]. The method consists of three key steps: computing cell state transition probabilities, inferring macrostates from the resulting Markov chain, and computing fate probabilities to these macrostates. For ease of comparison between other methods discussed here, our summary will focus mostly on the computation of transition probabilities.

CellRank computes a transition matrix from a $k$-NN graph using a combination of transcriptomic similarity and RNA velocity data. First, a $k$-NN graph is computed using cell transcriptomic similarities and then symmetrised. Edge weights are assigned based on similarity estimates between neighbouring cell states. The resulting graph is then converted into a matrix containing similarity estimates between neighbours. For each cell, transition probabilities are calculated from RNA velocity data by considering the correlation of the RNA velocity vector with displacement vectors corresponding to edges in the $k$-NN graph. These correlations are used to create a categorical distribution on the neighbours of the cell, giving transition probabilities. To better account for noise in the velocity data, the final transition probabilities are taken to be a linear combination of velocity-based probabilities and similarity-based probabilities.

We applied CellRank to the same 10,000 cell subset of the *Arabidopsis* dataset used for the StationaryOT and PBA analyses. Using the output transition matrix, we computed fate probabilities and assigned putative fates as previously described. In terms of putative fates, we found that CellRank matched 73% of cells to the atlas annotation, compared to 80% and 81% achieved by the StationaryOT analyses. The main differences occurred in the lateral root cap and xylem tissues (see Figure 15). CellRank also had less confidence in fate prediction, having an average fate probability of 45% compared to greater than 60% for all other methods. Finally, we note that CellRank uses a mixture of a directed transition matrix derived from RNA velocity and an undirected transition matrix computed from expression similarity. Therefore, it is not applicable to settings where velocity information is unavailable, such as scATAC-seq [4] and CyTOF [24] data, while StationaryOT and PBA can still be used.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(a)




(b)

Figure 16: Fate probabilities for (a) both StationaryOT methods compared to the annotation and (b) PBA and CellRank. The colour indicates the maximum fate probability (putative fate) of each cell and the colour saturation shows the magnitude of the fate probability.

## 2.9 Computing fates for large datasets

The running time for StationaryOT depends on the number of cells, with the main computational costs (in the case of the entropic regularisation) arising from (1) Sinkhorn iterations involving a series of matrix-vector products, and (2) solving a system of linear equations to compute fate probabilities. Computational cost therefore scales roughly quadratically in the number of cells (at least for a fixed number of iterations) and we found that datasets of up to $10^4$ cells could be processed directly using a straightforward implementation of the method. In order to compute cell fates for datasets with very large numbers of cells we propose two approaches.

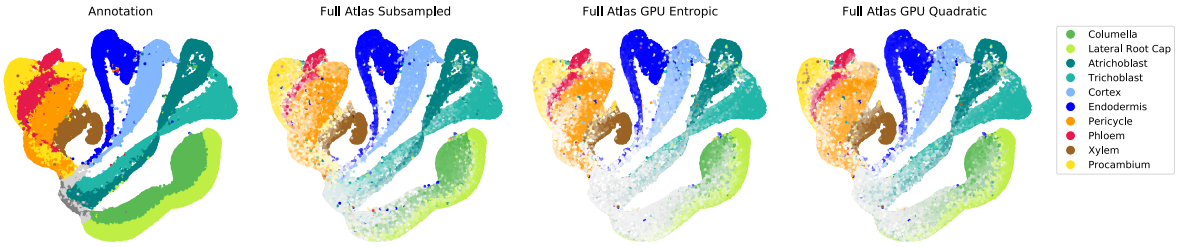



Figure 17: Atlas annotation on the full dataset ($1.1 \times 10^5$ cells) shown in UMAP coordinates compared to fate probabilities computed on the full dataset respectively using the subsampling approach (using entropic regularisation for each subproblem) and memory-efficient GPU implementations of StationaryOT with entropic and quadratic regularisations.

**Repeated subsampling** We first randomly partition the dataset of interest into $k$ subsets of size approximately $10^4$, or such that the computation time for StationaryOT is acceptable. Fates are computed for each subset, and this procedure is repeated $j$ times with repeated random partitioning. We then average the computed fates on a cell-by-cell basis, to produce aggregated fate probabilities.

We applied this approach to the full $1.1 \times 10^5$ cell atlas, partitioning it into 10 subsets and applying StationaryOT separately to each subset. This was repeated 10 times to account for sampling error.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Between the fates found directly for each subset and the consensus fates in the full atlas, 97% of cells shared the same putative fate and the maximum fate values had a correlation of 0.96. Accounting for all fate values, the correlation rose to 0.99.

**Memory-efficient GPU implementation with KeOps** For both entropic and quadratic regularisations, algorithms for solution of the optimal transport minimisation problem can be implemented using the KeOps library [5] so as to avoid storing all $N \times N$ matrices in memory. Along with GPU acceleration, this allows StationaryOT to be applied directly to datasets with many more cells than can be handled by the standard implementation due to memory constraints.

In the case of entropy-regularised optimal transport, the Sinkhorn algorithm can be implemented in a straightforward manner so that the transport plan (a matrix of size $N \times N$) is parameterised in closed form by two dual variables $u$ and $v$ (vectors each of length $N$) [25]. From here, one may construct a linear system to solve for fate probabilities of the form $Ax = b$ where $A$ is again parameterised by the dual variables $(u, v)$ and thus not explicitly stored in memory. For quadratically-regularised optimal transport, a similar representation of the problem in terms of dual variables holds. We provide an implementation of the semi-smooth Newton method proposed in [22, Algorithm 2] that utilises the KeOps library. As mentioned earlier in Section 2.4, quadratically regularised optimal transport has the property that the transport plans (and hence the transition laws of cells) will be sparse. In addition to being more interpretable, for large numbers of cells this means additional computational advantages, especially in terms of directly storing the entries of the sparse transition matrix, and computing the fate probabilities.

We applied StationaryOT to the full $1.1 \times 10^5$ cell *Arabidopsis* root dataset using the KeOps implementation with GPU acceleration, using both entropic and quadratic regularisations. We used the same parameter choices as used for the 10,000 cell subset in Section 2.8. We found that solution of the optimal transport problem took roughly 15 and 20 minutes using entropic and quadratic regularisations respectively. Computation of the fate probabilities for the entropic regularisation was significantly more time-consuming than for the quadratic regularisation, taking approximately 10 minutes and 2 minutes respectively. The difference in the runtimes reflects the fact that, compared to dense systems, sparse systems of linear equations can be solved much more efficiently using iterative methods. We compared the fates found for a 10,000 cell subset to the fates for StationaryOT with both entropic and quadratic regularisation on the full atlas and found them to perform similarly. For both methods, 90% of cells shared the same putative fate as the 10,000 subset and both had a 0.97 correlation for fate magnitudes accounting for all fates. With the entropic regularisation, the putative fate values were slightly higher correlated with the 10,000 cell subset, achieving a correlation of 0.92 compared to 0.87 for quadratic. All computations for the full atlas dataset were done on a Google Colaboratory instance with a 16GB NVIDIA Tesla V100 GPU.

# 3 Discussion

**Summary of our contributions** Optimal transport has been shown to be a widely applicable tool to the problem of trajectory inference in the setting where multiple time points are available [29, 18, 35, 10, 27, 7]. We demonstrate that optimal transport can be applied in a natural way to the stationary setting, where a single snapshot of a system at steady state is observed. The framework that we develop is theoretically justified and is naturally motivated by the Waddington's landscape analogy. Furthermore, our scheme boils down to a convex optimisation problem for which there are efficient and well-known methods of solution. The problem can also be generalised to incorporate additional information such as estimates of velocity. Motivated by these observations, we have developed a computational method which we call StationaryOT and show that it can scale to datasets of up to $10^5$ cells.

We demonstrate the efficacy of this method both on both real and simulated data. We find that in practice our method achieves similar performance to that of Population Balance Analysis (PBA) [41], but StationaryOT appears to be less sensitive to parameter choices and is capable of handling additional information such as velocity estimates. Since StationaryOT and PBA are methodologically distinct, the observation that both methods yield similar conclusions is strong evidence that the outputs reflect genuine biological signal, as opposed to artefacts of the methodology.

Overall, we have shown optimal transport to be a common framework for trajectory inference in the setting of both stationary snapshots and non-stationary, time-series data. This provides a unifying perspective for two problems that have traditionally been approached with separate methods.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Prospects for future work** In terms of future work, there are many potential avenues for extension of the present work. One major direction is the development of generative models, which can extrapolate information about the potential landscape beyond those cell states measured in experiment. We expect that the optimal transport perspective will be important for this, both conceptually and practically. Another relevant problem is that of examining the evolution of systems that are stationary on short timescales but nonstationary on large timescales – for instance, developmental biological systems such as haematopoiesis in humans are stationary on a fast timescale, but undergo changes on a slow timescale as individuals age. Finally, one could incorporate lineage-tracing to improve trajectory inference, as we have recently done in the non-stationary case [10].

## Software availability

An implementation of the StationaryOT computational method is available as an open-source software package at `https://github.com/zsteve/StationaryOT`.

## Acknowledgements

We would like to thank Rachel Shahan, Che-wei Hsu, and the Benfey Lab for their help in understanding the biological context of the *Arabidopsis* root. This work was supported in part by a UBC Affiliated Fellowship to S.Z., a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, an NFRF Exploration Grant, and a NSERC Discovery Grant to G.S.

## References

[1] Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, pages 1–7, 2020.

[2] Mathieu Blondel, Vivien Seguy, and Antoine Rolet. Smooth and sparse optimal transport. In *International Conference on Artificial Intelligence and Statistics*, pages 880–889. PMLR, 2018.

[3] Rowan D Brackston, Eszter Lakatos, and Michael PH Stumpf. Transition state characteristics during cell differentiation. *PLoS computational biology*, 14(9):e1006405, 2018.

[4] Jason D Buenrostro, Beijing Wu, Ulrike M Litzenburger, Dave Ruff, Michael L Gonzales, Michael P Snyder, Howard Y Chang, and William J Greenleaf. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490, 2015.

[5] Benjamin Charlier, Jean Feydy, Joan Alexis Glaunès, François-David Collin, and Ghislain Durif. Kernel operations on the gpu, with autodiff, without memory overflows. *arXiv preprint arXiv:2004.11127*, 2020.

[6] Yongxin Chen and Johan Karlsson. State tracking of linear ensembles via optimal mass transport. *IEEE Control Systems Letters*, 2(2):260–265, 2018.

[7] Karren Dai Yang, Karthik Damodaran, Saradha Venkatachalapathy, Ali C Soylemezoglu, GV Shivashankar, and Caroline Uhler. Predicting cell lineages using autoencoders and optimal transport. *PLoS computational biology*, 16(4):e1007828, 2020.

[8] Idan Efroni, Pui-Leng Ip, Tal Nawy, Alison Mello, and Kenneth D Birnbaum. Quantification of cell identity from single-cell gene expression profiles. *Genome Biology*, 16(9).

[9] David S Fischer, Anna K Fiedler, Eric M Kernfeld, Ryan MJ Genga, Aimée Bastidas-Ponce, Mostafa Bakhti, Heiko Lickert, Jan Hasenauer, Rene Maehr, and Fabian J Theis. Inferring population dynamics from single-cell rna-sequencing time series data. *Nature biotechnology*, 37(4):461–468, 2019.

[10] Aden Forrow and Geoffrey Schiebinger. A unified framework for lineage tracing and trajectory inference. *bioRxiv*, 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[11] Gunsagar S. Gulati, Shaheen S. Sikandar, Daniel J. Wesche, Anoop Manjunath, Anjan Bharadwaj, Mark J. Berger, Francisco Ilagan, Angera H. Kuo, Robert W. Hsieh, Shang Cai, Maider Zabala, Ferenc A. Scheeren, Neethan A. Lobo, Dalong Qian, Feiqiao B. Yu, Frederick M. Dirbas, Michael F. Clarke, and Aaron M. Newman. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science*, 367(6476):405–411, 2020.

[12] Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan Karlsson. Estimating ensemble flows on a hidden markov chain. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1331–1338. IEEE, 2019.

[13] Laleh Haghverdi, Maren Büttner, F Alexander Wolf, Florian Buettner, and Fabian J Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10):845, 2016.

[14] Helge Holden, Kenneth H Karlsen, and Knut-Andreas Lie. *Splitting methods for partial differential equations with rough solutions: Analysis and MATLAB programs*, volume 11. European Mathematical Society, 2010.

[15] Allon M Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5):1187–1201, 2015.

[16] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018.

[17] Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B. Schiller, Dana Pe'er, and Fabian J. Theis. Cellrank for directed single-cell fate mapping. *bioRxiv*, 2020.

[18] Hugo Lavenant, Stephen Zhang, Young-Heon Kim, and Geoffrey Schiebinger. Towards a mathematical theory of trajectory inference.

[19] Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

[20] Chieh Lin and Ziv Bar-Joseph. Continuous-state hmms for modeling time-series single-cell rna-seq data. *Bioinformatics*, 35(22):4707–4715, 2019.

[21] Dawn Lin, Andrey Kan, Jerry Gao, Edmund Crampin, Philip D Hodgkin, and Shalin H Naik. Assessment of clonal kinetics reveals multiple trajectories of dendritic cell development. *bioRxiv*, page 167635, 2017.

[22] Dirk A Lorenz, Paul Manns, and Christian Meyer. Quadratically regularized optimal transport. *Applied Mathematics & Optimization*, pages 1–31, 2019.

[23] Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

[24] Subarna Palit, Christoph Heuser, Gustavo P de Almeida, Fabian J Theis, and Christina E Zielinski. Meeting the challenges of high-dimensional single-cell data analysis in immunology. *Frontiers in immunology*, 10:1515, 2019.

[25] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[26] Plant Illustrations. Root illustrations, 2017.

[27] Neha Prasad, Karren Yang, and Caroline Uhler. Optimal transport using gans for lineage tracing. *arXiv preprint arXiv:2007.12098*, 2020.

[28] Rahim Rahni and Kenneth D Birnbaum. Week-long imaging of cell divisions in the arabidopsis root meristem. *Plant Methods*, 15(30), 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[29] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

[30] Rachel Shahan, Che-Wei Hsu, Trevor M Nolan, Benjamin J Cole, Isaiah W Taylor, Anna Hendrika Cornelia Vlot, Philip N Benfey, and Uwe Ohler. A single cell arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants. *bioRxiv*, 2020.

[31] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):1–16, 2018.

[32] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M 3rd Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902, 2019.

[33] Patrick S Stumpf, Rosanna CG Smith, Michael Lenz, Andreas Schuppert, Franz-Josef Müller, Ann Babtie, Thalia E Chan, Michael PH Stumpf, Colin P Please, Sam D Howison, et al. Stem cell differentiation as a non-markov stochastic process. *Cell Systems*, 5(3):268–282, 2017.

[34] Anders Tolver. An introduction to markov chains. *Department of Mathematical Sciences, University of Copenhagen*, 2016.

[35] Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. *arXiv preprint arXiv:2002.04461*, 2020.

[36] Thinh N Tran and Gary D Bader. Tempora: Cell trajectory inference using time-series single-cell rna sequencing data. *PLoS computational biology*, 15(30):e1008205, 2020.

[37] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, 32(4):381, 2014.

[38] Sophie Tritschler, Maren Büttner, David S Fischer, Marius Lange, Volker Bergen, Heiko Lickert, and Fabian J Theis. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development*, 146(12):dev170506, 2019.

[39] Conrad Waddington. H. 1957. the strategy of the genes, 1959.

[40] Ping Wang, Yidong Chen, Jun Yong, Yueli Cui, Rui Wang, Lu Wen, Jie Qiao, and Fuchou Tang. Dissecting the global dynamic molecular profiles of human fetal kidney development by single-cell rna sequencing. *Cell reports*, 24(13):3554–3567, 2018.

[41] Caleb Weinreb, Samuel Wolock, Betsabeh K Tusi, Merav Socolovsky, and Allon M Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.

[42] F Alexander Wolf, Fiona K Hamey, Mireya Plass, Jordi Solana, Joakim S Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J Theis. Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*, 20(1):1–9, 2019.

[43] Scott A Yuzwa, Michael J Borrett, Brendan T Innes, Anastassia Voronova, Troy Ketela, David R Kaplan, Gary D Bader, and Freda D Miller. Developmental emergence of adult neural stem cells as revealed by single-cell transcriptional profiling. *Cell reports*, 21(13):3970–3986, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# A Computation of fate probabilities and conditional MFPT

**Fate probabilities** As in the main text, we consider a discrete state space $\overline{\mathcal{X}} = \{x_i\}_{i=1}^N$ and take $P$ to be a matrix of transition probabilities. For brevity, let $\overline{\mathcal{X}}_\varnothing$ denote the set of sink states for which we want to compute fate (absorption) probabilities. Then we compute fate probabilities via the fundamental matrix as follows [34, Theorem 28]:

- Make all sink states absorbing, i.e. produce a modified transition matrix $\tilde{P}$ from $P$ such that for each sink $x_i \in \overline{\mathcal{X}}_\varnothing$, we have
$$\tilde{P}_{ij} = \mathbf{1}_{i=j}.$$
- We reorder the rows and columns of $\tilde{P}$ to form the matrix
$$\begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}, \tag{20}$$
where $Q$ corresponds to transitions between transient states and $R$ corresponds to transitions from transient states to sink states.
- We next compute the fundamental matrix of the absorbing Markov chain as
$$N = (I - Q)^{-1}. \tag{21}$$
- From the fundamental matrix, we compute $B = NR$. The entry $B_{ij}$ is the absorption probability for sink $j$ for a chain starting from state $i$, i.e.
$$B_{ij} = \mathbb{P}[X_\infty = x_j | X_0 = x_i].$$

**Numerical considerations** In practice we have found that the computation of fate probabilities can sometimes suffer from numerical issues that result in the fate probability problem being ill-conditioned. This typically appears to be the case for the quadratic regularisation when the regularisation parameter $\varepsilon$ is chosen to be small, resulting in a disconnected graph or degenerate transition laws for outlier cells due to the sparsity of the resulting transport plan. In practice, one can always regularise transition matrices before computing fate probabilities by mixing in a small undirected diffusion component, as is done in [17], i.e.
$$P = (1 - f)P_{\text{statOT}} + fP_{\text{diffusion}},$$
where $f$ is chosen to be small (e.g. $f = 0.05$), and $P_{\text{diffusion}}$ is a matrix of undirected transition probabilities, etc.
$$(P_{\text{diffusion}})_{ij} \propto \exp\left(-\frac{\|x_i - x_j\|^2}{h}\right),$$
and $h$ is a user-specified bandwidth parameter.

**Conditional MFPT** The conditional mean first passage time (MFPT) from $x_i$ to $x_j$ is defined as
$$\mathbb{E}[t | X_0 = x_i, X_t = x_j].$$
To compute the conditional MFPT, we follow the steps laid out in [41, Theory Supplement, Section 3.3].

**Ground truth fate probabilities and ground truth simulation time** In simulations, to compute the ground truth fate probability of a state $x_i$, we simulated 100 trajectories from start to finish, initialised at $x_i$ when $t = 0$. The fate probability of $x_i$ to fate $j$ was taken to be the proportion of simulated trajectories terminating in fate $j$.

For the ground truth simulation time, we sampled $10^4$ trajectories $X_t^{(i)}$ from the underlying process. For each trajectory, we also recorded the simulation time of each state and aggregated these into a reference table. Then, for each sampled state $x_i \in \overline{\mathcal{X}}$, we took the corresponding ground truth simulation time $t_i$ to be the average simulation time of the 250 nearest points in the reference.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# B Root atlas application details

## B.1 Data preprocessing

The *Arabidopsis thaliana* atlas contains 16 replicates that were preprocessed using the COPILOT pipeline [30]. The replicates were integrated using the Seurat integration pipeline [30, 32]. The resulting atlas was annotated by Shahan et al. with cell lineages and developmental zones using an ensemble of correlation to published expression profiles, Index of Cell Identity scoring, and expression of known marker genes [30, 8]. An ensemble pseudotime was also calculated from RNA velocity using scVelo and based on transcriptional diversity using CytoTRACE [11, 1].

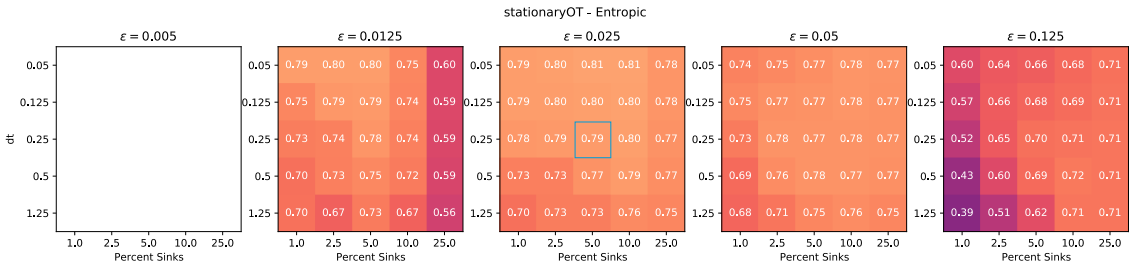
Growth rates were calculated from imaging data of the *Arabidopsis* meristem by averaging the division time observed for all cells in a lineage and converting the division times to an average daily growth rate [28]. In two cases, the imaging data grouped multiple lineages from our data into a single category. In these cases, we used the average growth rate for all cells in that category for each lineage.

## B.2 CellRank

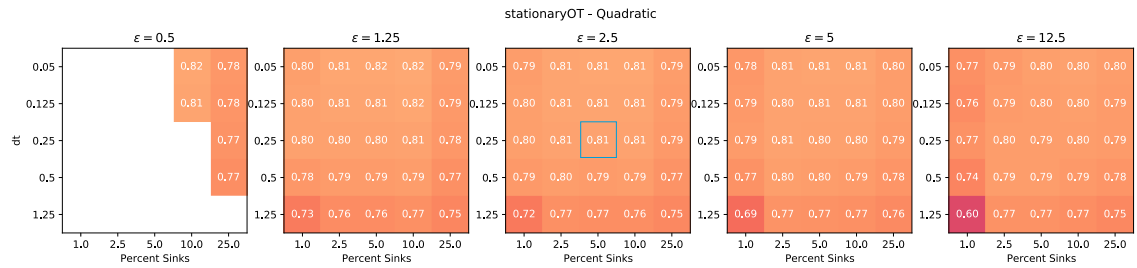
Based on a parameter sweep, we chose $k = 10$ for the $k$-NN graph, 0.5 for the weight on the transcriptional similarity kernel, and a softmax of 7.5 (See Figure 19c). Using these settings, CellRank was unable to automatically detect all terminal and initial macrostates (see Figure 24). Therefore, macrostates were assigned manually to the same sinks used for PBA and StationaryOT. Sources were defined to be all cells in the stem cell niche and the putative quiescent centre.

## B.3 Parameter variation

We chose the parameters shown in the main text by performing an initial parameter sweep around a 6-hour time step ($\Delta t = 0.25$) and 5% sinks. The number of sinks was chosen based on the 6-hour time step, the growth rates, and the assumption that the root is in equilibrium so that that number of sinks balanced with the number of cells that would be added by growth. We evaluated the runs based on the percentage of putative fates that matched the annotation. To test the robustness of each method, we tested a grid of parameters taking all combinations of changes of a factor of two and five in each direction.




(a)



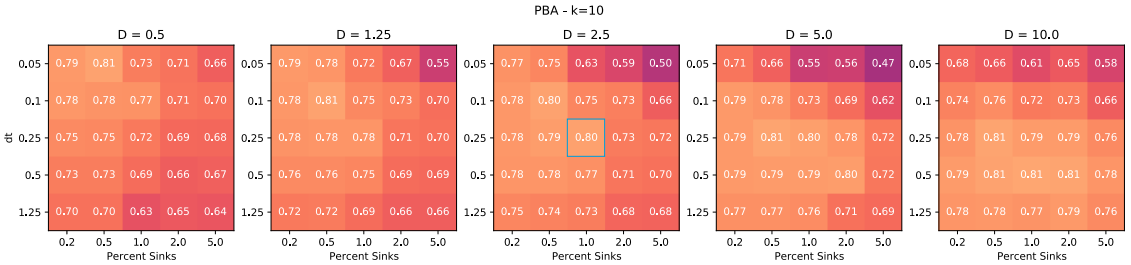

(b)

Figure 18: Parameter variation for StationaryOT with (a) entropic and (b) quadratic regularisation. The value for each combination represents the percentage of putative fates matching the annotation. Blank squares indicate runs where fates could not be calculated due to underflow or a singular coupling. The run shown in the main text is indicated in blue.

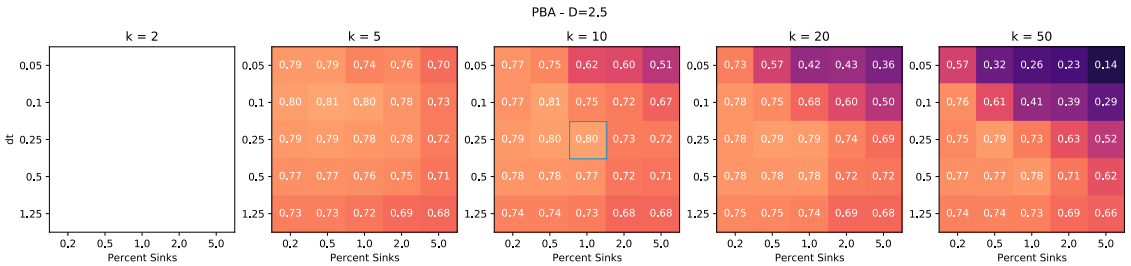
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For StationaryOT with entropic regularisation and small values of $\varepsilon$, we added a small amount of Gaussian noise to the coupling to avoid a few outlier cells remaining stationary. As shown in Figure 18a, the match to the annotation changed no more than 2% when changing a parameter by a factor of two, but performance significantly degraded when multiple factors were off by a factor of two or more. StationaryOT with quadratic regularisation was more robust to parameter changes, differing by no more than 2% for a factor of two change in a single parameter and remaining within 2% for many combinations where multiple parameters were off by a factor of five. When using quadratic regularisation with very small values of $\varepsilon$, the coupling matrix becomes singular and we cannot calculate fate probabilities. When using entropic regularisation for very small values of epsilon, the coupling could not be computed due to underflow. For both of these issues, runs are shown as blank squares in Figure 18b.

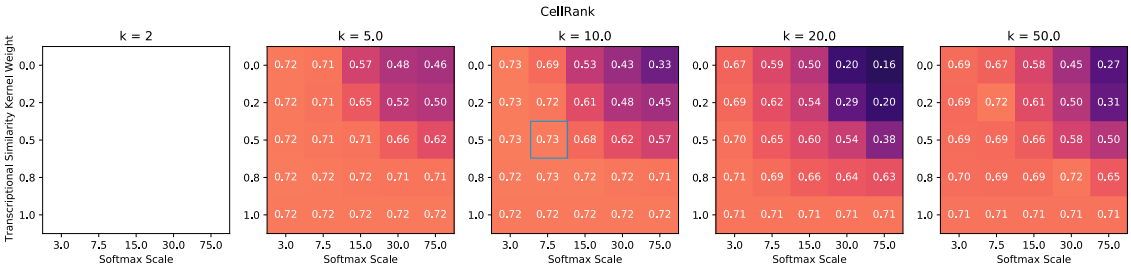
Like StationaryOT, PBA has parameters for the number of sinks and time step. However, instead of using a single regularisation parameter, PBA has both diffusivity ($D$) and number of neighbours in the $k$-NN graph ($k$). To limit the number of runs to 256, $D$ and $k$ were fixed while the other three parameters were varied (see Figure 19a-19b). We found that PBA was highly sensitive to the number of sinks in addition to the total flux. Using our estimate that the root would stay in equilibrium when 5% of cells were set as sinks for a 6-hour time step, around 70% of cells matched the annotation. Reducing the number of sinks to 1% for a 6-hour interval increased the match to 80%. However, in many applications the annotation would not be available to evaluate the runs, and the best combination of parameters could be missed.




(a)




(b)




(c)

Figure 19: Parameter variation for PBA (a-b) and CellRank (c), evaluated by the percentage of cells whose putative fate matched the annotation. For both PBA and CellRank, the graphs became disconnected when only two neighbours were used and thus fate probabilities could not be computed. The run shown in the main text is indicated in blue.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## B.4 Supplemental figures

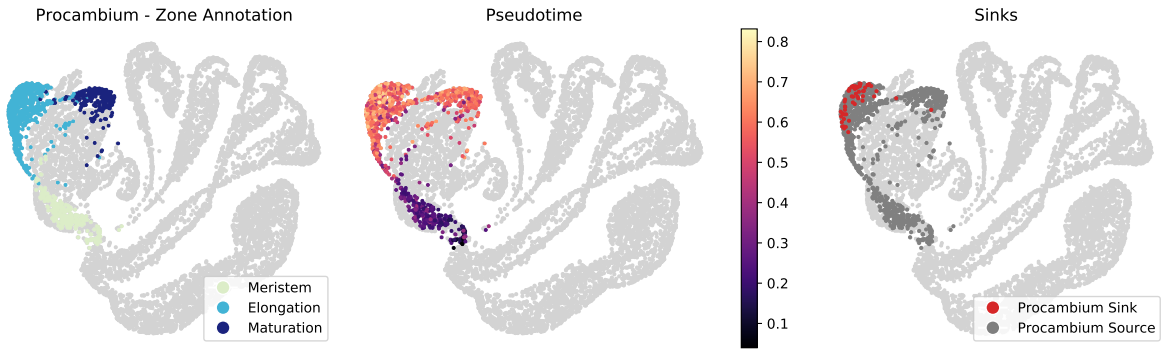



Figure 20: All four methods poorly matched the annotation for maturation procambium cells (see Figure 15). We believe this occurred due to a disagreement between the pseudotime and zone annotations, where procambium cells in the elongation zone were given a higher pseudotime than those in the maturation zone, resulting in cells from the elongation zone incorrectly being set as terminal states.

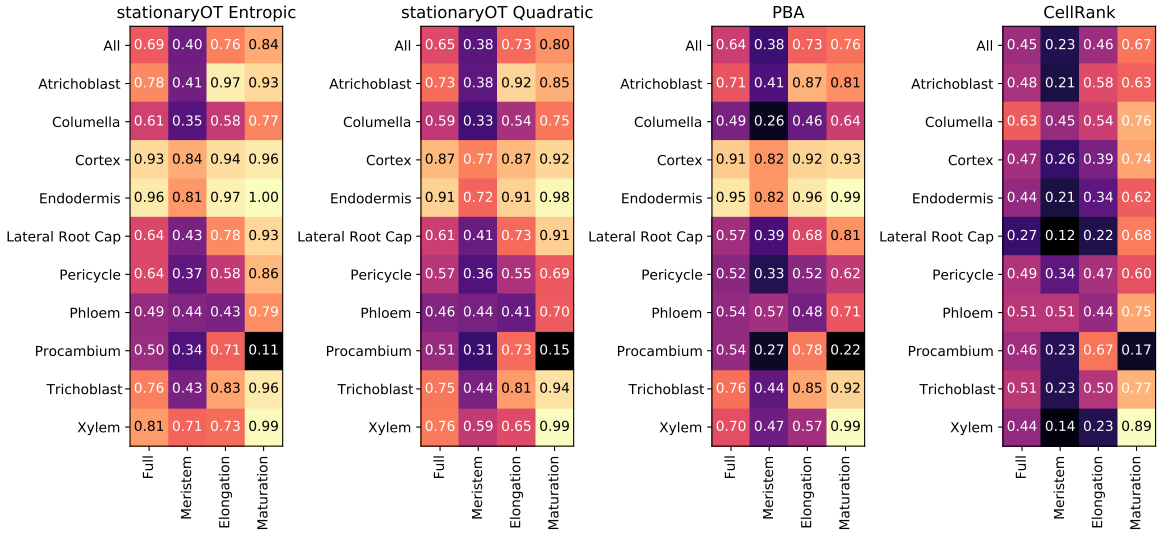



Figure 21: Average fate probabilities matching the annotation by cell type and zone for both StationaryOT methods, PBA, and CellRank.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.02.433630; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 23: Lineage annotation compared to cell fate probabilities for both StationaryOT methods, PBA, and CellRank.

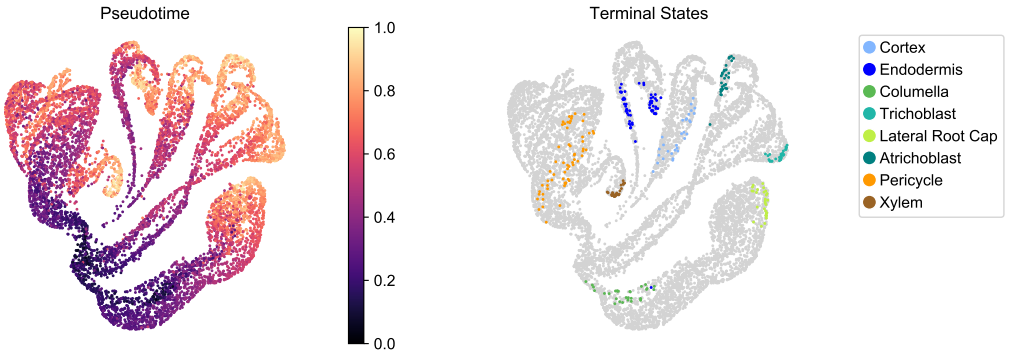



Figure 24: Terminal states found using automatic detection functionality offered by the CellRank package, coloured by their corresponding lineage (right). No terminal states were identified for the phloem and procambium lineages. Additionally, as is clear from pseudotime (left), some states that are intermediate are miss-classified as terminal.