Inferring cell-specific causal regulatory networks from drift and diffusion

Stephen Y Zhang [1] **Michael P H Stumpf** [1]

Abstract

Cellular dynamics are fundamentally governed by networks of interacting genes, and inferring these interactions from data is a central problem in systems biology. Gene expression profiling at single-cell resolution is now routine, allowing for heterogeneity of cell state to be studied at scale. In contrast, the vast majority of network inference methods work at the population level to construct a static gene regulatory network, and thus do not allow for inference of differential regulation across sub-populations. Building upon recent inference methods that model single cell dynamics using drift-diffusion processes, we propose a method for inferring cell-specific causal networks and show some preliminary results.

1. Introduction

Identifying and characterizing gene regulatory interactions is one of the end goals of modern transcriptomic studies. Over the last decade, molecular profiling at single cell resolution has become standard. An oft-cited advantage of single cell assays is that heterogeneous population structure is preserved, allowing for the detection and study of rare cellular phenotypes that would be unobservable from bulk studies (Buettner et al., 2015). The single cell "revolution" has spurred rapid advances in data analysis and modelling methods such as trajectory inference (Tritschler et al., 2019), dimensionality reduction and data integration. However, to date the majority of gene regulatory network inference methods aim to reconstruct a static network (Pratapa et al., 2020) that describes the set of possible interactions occurring within an observed population. Given the biological importance of cellular heterogeneity, a natural expectation is that variations in transcriptional state may correspond to variations in (cell state dependent) regulatory interactions which cannot be represented as static networks (Stumpf, 2021).

Recent works have dealt with cell-specific networks, where multiple networks are learned from a single-cell population, thus allowing for variations in learned regulatory relationships across a dataset (Akers & Murali, 2021). Such methods leverage neighbourhood or cluster information to construct networks that are cell- or context-specific. To our knowledge, existing works have focused on inferring undirected networks (Zhang et al., 2022; Wang et al., 2021). If we hope to infer *causal* interactions, then information about underlying dynamics in a cell population is essential. The problem of estimating dynamics from data is a well studied problem, and we appeal to the corresponding literature (Teschendorff & Feinberg, 2021; Ding et al., 2022).

2. Method

Since we aim at inferring cell-specific networks, it is necessary to avoid coarse-graining of cell states as much as possible. Hence, we will avoid trajectory inference methods that require clustering or smoothing of cell states. We restrict our focus to a class of methods that model the dynamics as a Markov, drift-diffusion process over a manifold of cell states (Weinreb et al., 2018; Lange et al., 2022). Let $\mathcal{X} = \{x_i\}_{i=1}^N$ be a sample of a population of cells, where $x_i \in \mathbb{R}^M$ is the vector of mRNA expression values for cell i. Let $\mathcal{M}$ be a graph constructed from $\mathcal{X}$ that approximates some underlying manifold of cell states (Moon et al., 2018). We assume we have access to a transition matrix $P \in \mathbb{R}^{N\times N}$ that is supported on $\mathcal{M}$ and adequately describes the "true" dynamics of the biological process on the manifold (in practice, as we do in Section 3, one could use a variety methods in the recent literature to estimate P). Then $\mathcal{M}$ equipped with P encodes a discrete Markov chain that approximates the dynamical process in some sense. In what follows, we will write $[\![n]\!] = \{1, \dots, n\}$ for $n \in \mathbb{N}$.

Restricted Directed Information For any cell $x \in \mathcal{M}$, we can construct (e.g. by a truncated Gaussian) a probability distribution supported on its neighbourhood, π_0^x. One may then consider a Markov process $X_t, t \geqslant 0$ started from $X_0 \sim \pi_0^x$ and evolving under P. For some fixed t, the resulting *coupling* is

$$(X_0, X_t) \sim \mathrm{diag}(\pi_0^x)P^t =: \gamma_t^x. \tag{1}$$

[1]Mathematics and Statistics, University of Melbourne, Australia. Correspondence to: Stephen Zhang <stephenz@student.unimelb.edu.au>.

The 2022 ICML Workshop on Computational Biology. Baltimore, Maryland, USA, 2022.

Denote by a pair of genes (j, k) the causal relationship $j \to k$. We aim to quantify the restricted directed information (RDI) (Rahimzamani & Kannan, 2016) from knowledge of the coupling γ_t^x and gene expression states:

$$\mathrm{RDI}(X_0[j] \to X_t[k]) := I(X_0[j], X_t[k] | X_0[k]), \quad (2)$$

where $I(X, Y|Z) := \mathbf{E}_{X,Y,Z}\left[\log \rho_{XY|Z} - \log(\rho_{X|Z} \otimes \rho_{Y|Z})\right]$ is the conditional mutual information, and $X_t[j]$ is the expression of gene j in cell X_t. Doing this for all cells $x_i \in \mathcal{M}$ and gene pairs $j, k \in [\![M]\!]$, we arrive at a tensor $\hat{\boldsymbol{G}} \in \mathbb{R}^{N \times M \times M}$ such that $\hat{\boldsymbol{G}}_{ijk}$ is the RDI score of $j \to k$ in cell x_i.

Since the RDI score matrix $\hat{\boldsymbol{G}}_i$ for each cell is learned from a relatively small local neighbourhood, it is potentially extremely noisy. Furthermore, it is impractical (due to limited data and computational resources) to condition on potential confounders (Qiu et al., 2020). Thus, RDI scores may contain noise and undesirable signal resulting from indirect interactions. To deal with this, we first filter interaction scores using context likelihood of relatedness (CLR) (Madar et al., 2010) and then solve a manifold-regularized optimization problem to both denoise the interaction signals and attenuate signal arising from indirect interactions (Qiu et al., 2020; Zhang et al., 2022).

Context likelihood of relatedness. Given a (generally dense) matrix of RDI scores $\hat{G}$, we employ the context likelihood of relatedness (CLR) algorithm (Madar et al., 2010) to produce a filtered matrix $\tilde{G}$ as follows. For each pair of genes (i, j), compute z_i (resp. z_j) to be the z-score of G_{ij} with respect to $G_{i\cdot}$ (resp. $G_{\cdot j}$). Then we define the CLR score for (i, j) to be

$$\tilde{G}_{ij} = \frac{\hat{G}_{ij}}{2}\sqrt{\max(0, z_i)^2 + \max(0, z_j)^2}. \quad (3)$$

Applying CLR filtering along the first axis of $\hat{\boldsymbol{G}}$, we obtain the filtered tensor $\tilde{\boldsymbol{G}}$. We remark that we have made a modification to the original approach of (Madar et al., 2010) here, namely that we weight the CLR score by the initial MI value $\hat{G}_{ij}$. This is important since CLR was originally designed to filter interactions in static networks. In the setting of cell-specific networks, very few edges may be "active" in a given cell's context, and so entire rows or columns of $\hat{G}$ may consist only of noise. Computing z-scores along those rows or columns would put both noise and signal on the same scale, hence we choose to weight by the original value.

Smoothing and denoising The tensor $\tilde{\boldsymbol{G}}$ contains a noisy matrix of interaction scores, one for each measured cell. For notational convenience we will matricize $\tilde{\boldsymbol{G}}$ to form $\tilde{G} \in \mathbb{R}^{N \times M^2}$, i.e. each row is a length-M^2 unfolded score matrix. We propose to solve the optimization problem $G = \arg\min_X \mathcal{L}(X; L, \tilde{G})$ where

$$\mathcal{L}(X; L, \tilde{G}) = \frac{1}{2}\sum_{i=1}^{N} w_i \|X_{i\cdot} - \hat{G}_{i\cdot}\|_2^2 + \frac{\lambda_1}{2}\operatorname{tr}(X^\top L X) + \lambda_2 \sum_{i=1}^{N} w_i \|X_{i\cdot}\|_1, \quad (4)$$

λ_1, λ_2 are hyperparameters that control regularization strength, and L is the graph Laplacian of $\mathcal{M}$. The term associated to λ_1 is a Laplacian regularization corresponding to an assumption that regulatory relationships should vary smoothly with changes in cell state, i.e. $\operatorname{tr}(X^\top L X)$ is large for rapidly fluctuating X. The term associated to λ_2 is a Lasso term that encourages X to be sparse. This is a standard assumption about the nature of biological networks (Zhang et al., 2022). Together, the objective (4) encourages both parsimony and sharing of information along the manifold. The problem is a case of L1-L2-regularized least squares, and can be solved using a simple alternating scheme which we describe in the appendix.

Using a backward transition matrix Since in the above we consider couplings of the process at times $(0, t)$, estimates of the MI scores will be biased by some time lag. In order to remedy this, we must instead consider couplings for times $(-t, t)$. Rigorously speaking, time-reversal of a diffusion-drift process away from equilibrium is not always well-posed. In practice however, backward operators in some sense have been constructed by transposing the transition matrix (Lange et al., 2022; Li et al., 2020) (see Appendix A.1 for further discussion). Given an approximate backward operator Q, a time-symmetric coupling $\gamma_t^x = (Q^\top)^t \operatorname{diag}(\pi_0^x) P^t$ can be constructed.

3. Results

Overview We now discuss an example in which we seek to infer cell-specific networks from simulated single-cell data. In particular, we are interested in the scenario where a single, static network does not satisfactorily describe the true interactions. Consider the network in Figure 1(a, b), where a bifurcation driven by a toggle-switch feeds into one of two "modules". Modules A and B involve the same genes $\{6, 7, 8, 9\}$, but with completely different interactions between the species: the flow of information in each module is the mirror image of the other.

A static network (in the graphical sense) is insufficient to describe this system, since the directionality of some interactions may be dependent on the context. Consider gene 7, which may either be activated by gene 6 (in module A) or by gene 8 (in module B). This could conceivably be understood to be a *hyperedge*, since a corresponding Boolean rule

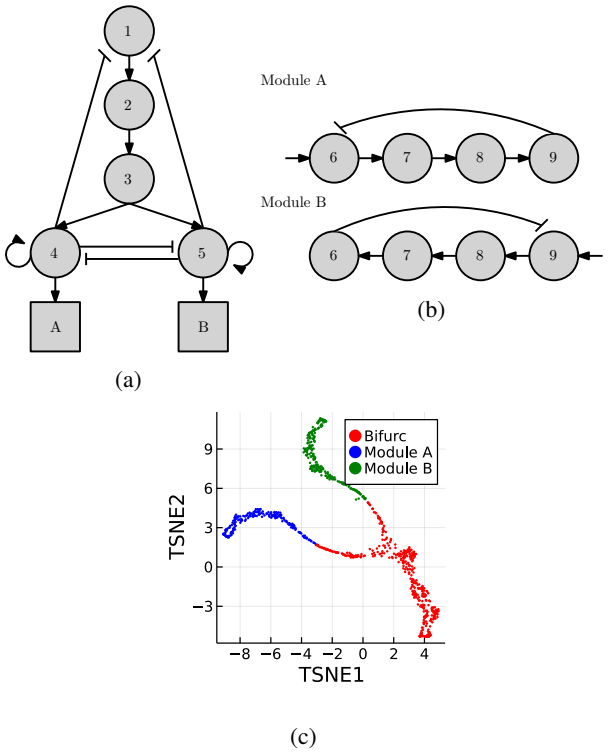


Figure 1. **Example network.** (a) 5-gene bifurcation structure feeding into two modules A, B involving the same 4 genes. (b) 4-gene modules A, B each involving different regulatory interactions. (c) tSNE embedding of simulated cell profiles, colored by cluster.

would be

$$g7 \leftarrow (g6 \wedge (g4 \wedge \neg g5)) \vee (g8 \wedge (\neg g4 \wedge g5)).$$

Short of attempting to perform rigorous inference for higher-order interactions, cell-specific networks allow us to disentangle interactions for a certain subclass of higher-order interactions: those which can be understood to be locally first-order, conditional on some latent variable that is encoded in the cellular *context*. Finally, since existing cell-specific network inference methods (Zhang et al., 2022; Wang et al., 2021) infer undirected edges only, such methods are unable to distinguish the direction of causality in this example.

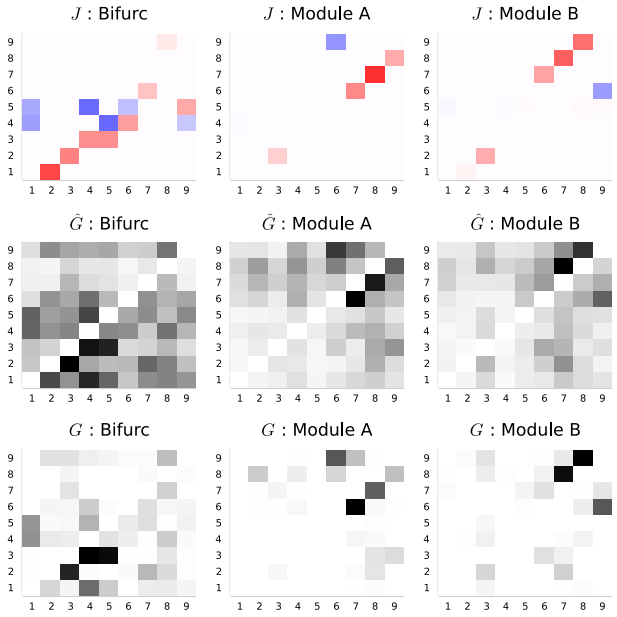
Simulation We simulated the network of Figure 1 using the BoolODE software package (Pratapa et al., 2020). This uses a chemical Langevin scheme for simulating biologically plausible expression dynamics: $\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}B_t$ where f, g have specific forms dictated by the reaction network. From 1,000 independent realizations of the system, we produced a set of 1,000 sampled cells by sampling a cell state from each trajectory at a time chosen uniformly at random. As a measure of ground truth interactions for each cell x_i, we computed the corresponding Jacobians $J_{ijk} = \partial_j f_k(x_i)$.

Inferring dynamics As an input, our method requires a cell-state transition matrix P. This can be estimated using a number of methods in practice. In our case, we construct P from transcriptomic velocity ("RNA velocity") measurements as follows. For each observed cell x_i, we calculate the corresponding velocity vector $v_i = f(x_i)$. From this, we may write $P_{ij} = Z^{-1} \exp(\langle x_j - x_i, v_i \rangle / \sigma) \mathbf{1}\{j \in \mathcal{N}[i]\}$, where Z is the appropriate normalizing constant and $\mathbf{1}$ denotes a 0-1 indicator. We set $\mathcal{N}$ to be the k-NN neighbourhood with $k = 25$, and took $\sigma = 0.5$ (Li et al., 2020).

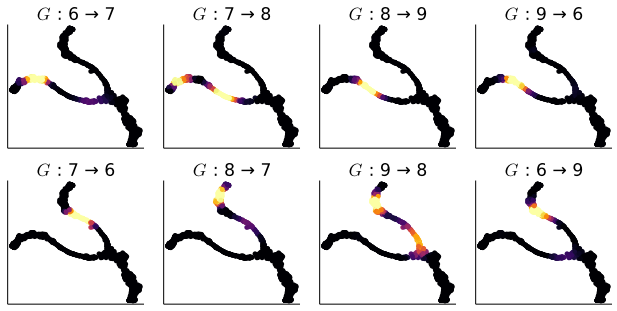
Recovering causality We applied the method of Section 2 with a backward matrix obtained via the transpose (see Section A.1), for $t = 3$, $\lambda_1 = 25.0, \lambda_2 = 0.025$. For construction of the neighbourhood π_0^x we used quadratic optimal transport (Matsumoto et al., 2022), which produces a local neighbourhood density similar to a truncated Gaussian. In Figure 2(a), we show the cell-specific networks averaged over the clusters delineated in Figure 1(c). We observe that the the denoised result $\boldsymbol{G}$ resembles the ground truth. Furthermore, the averaged networks for module A and module B are directed and reflect the mirrored casual relationships described in Figure 1(b). Averaging networks is convenient for summarizing the inference output for cell-specific networks, but in Figure 2(b) we show the interaction scores for a subset of edges on genes 6-9. It is clear visually that reversing the direction of each edge corresponds to switching branches in the tSNE embedding. Our method is able also to resolve the temporal behaviour of some interactions. For instance, the interaction $7 \to 8$ is active along its branch first when expression of gene 7 leads to activation of gene 8, and later when gene 7 is switched off, leading to gene 8 being switched off.

Cell-specific edge detection Finally, we consider the general problem of cell-specific edge detection. For each triple (i, j, k) with $i \in [\![N]\!], j, k \in [\![M]\!]$ we treat the problem of detecting in cell x_i an edge $j \to k$ as a binary classification problem with some threshold q. To construct the set of true positives, we consider the matrix ΠJ, where Π is a neighbourhood transition matrix such that $\Pi_{i\cdot} = \pi_0^{x_i}$. The motivation for considering this instead of simply J is as follows: cell-specific interactions are necessarily inferred by leveraging neighbourhood information, and so strict cell-wise comparison to a ground truth would be overly stringent and sensitive to small perturbations in the expression space. By computing RJ, the ground truth signal is smoothed over the same neighbourhood allowing for more robust assessment of classifiers. Subsequently, we classified all edges where $(RJ)_{ijk} > 0.5$ to be a "true" edge, and the fraction of edges was 4%.

In Figure 3, we show the Precision-Recall (PR) curves calculated from the raw RDI scores $\hat{\boldsymbol{G}}$ and the smoothed, de-



(a)



(b)

Figure 2. (a) Top: true interactions averaged by cluster; Middle: $\hat{\boldsymbol{G}}$ averaged by cluster; Bottom: $\boldsymbol{G}$ averaged by cluster.(b) Interactions corresponding to modules A and B shown for each cell on the tSNE embedding.

noised scores $\boldsymbol{G}$ respectively. In addition to the velocity kernel considered previously, we also considered kernels constructed using optimal transport (Zhang et al., 2021), population balance analysis (Weinreb et al., 2018) and diffusion pseudotime (Haghverdi et al., 2016). Furthermore, we considered the following baselines: reversed velocity (this should result in reversal of causality), undirected (pure diffusion, resulting in loss of temporal information), and random predictor (edges predicted at random with probability 0.04). See Appendix A.3 for hyperparameter choices. Comparing the AUPR for $\hat{\boldsymbol{G}}$ to those for $\boldsymbol{G}$, it is immediately clear that the filtering, smoothing and denoising steps result in an increased AUPR, i.e. improved detection ability for cell-specific edges. This can be understood as a combined effect of (1) removal of indirect interactions using the CLR algorithm, and (2) sharing of information

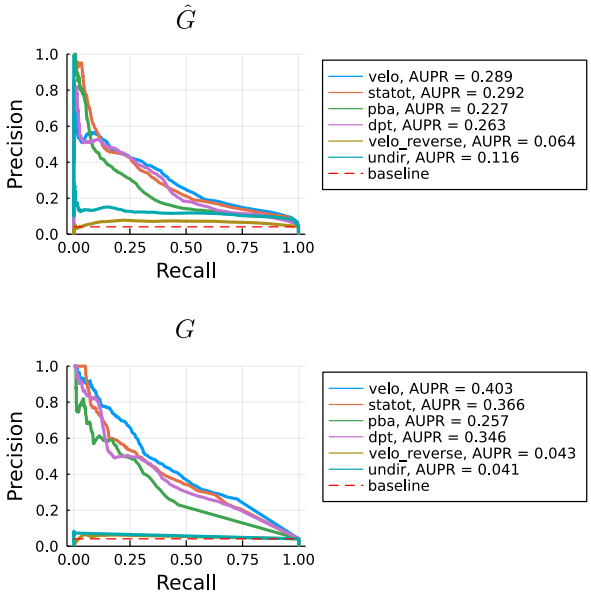


Figure 3. Precision-Recall (PR) curves for cell-specific edge detection from $\hat{\boldsymbol{G}}$ (raw RDI values) or $\boldsymbol{G}$ (smoothed and denoised) various choices of transition kernel reconstruction. Area under PR curve (AUPR) is shown.

along the cellular manifold and attenuation of noise using the smooth and sparse optimisation. The velocity kernel performs best (AUPR 0.403), followed by StationaryOT (AUPR 0.366). This is in keeping with the observation of (Qiu et al., 2020) that velocity estimates contain information about the true temporal coupling of cells. Reassuringly, both the undirected kernel and reversed velocity kernel perform very poorly (AUPR 0.041, 0.043 respectively). This reflects the fact that the kernel must accurately encode temporal information if causal interactions are to be inferred.

4. Conclusion

Motivated by recent developments in dynamical inference for single-cell data, we have presented a framework which enables reconstruction of cell-specific, causal networks. Given a transition kernel that approximates underlying cellular dynamics, our method relies on computing the Restricted Directed Information (RDI) (Rahimzamani & Kannan, 2016) for pairs of genes to measure causal relationships and makes no assumptions on linearity of interactions nor the distribution of gene expression counts. We plan to next investigate methods for filtering cell-specific networks to remove indirect edges, and regression methods for sharing information across the cellular manifold. We will also investigate other measures of causality, such as Granger causality (Qiu et al., 2020), with which one might hope to infer edges that are both directed and signed.

References

Akers, K. and Murali, T. Gene regulatory network inference in single-cell biology. *Current Opinion in Systems Biology*, 26:87–97, 2021.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., and Stegle, O. Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. *Nature biotechnology*, 33 (2):155–160, 2015.

Ding, J., Sharon, N., and Bar-Joseph, Z. Temporal modelling using single-cell transcriptomics. *Nature Reviews Genetics*, pp. 1–14, 2022.

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., and Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10):845–848, 2016.

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H. B., et al. Cellrank for directed single-cell fate mapping. *Nature methods*, pp. 1–12, 2022.

Li, T., Shi, J., Wu, Y., and Zhou, P. On the mathematics of rna velocity i: theoretical analysis. *bioRxiv*, 2020.

Madar, A., Greenfield, A., Vanden-Eijnden, E., and Bonneau, R. Dream3: network inference using dynamic context likelihood of relatedness and the inferelator. *PloS one*, 5(3):e9803, 2010.

Matsumoto, T., Zhang, S., and Schiebinger, G. Beyond $k$nn: Adaptive, sparse neighborhood graphs via optimal transport. Submitted, 2022.

Moon, K. R., Stanley III, J. S., Burkhardt, D., van Dijk, D., Wolf, G., and Krishnaswamy, S. Manifold learning-based methods for analyzing single-cell rna-sequencing data. *Current Opinion in Systems Biology*, 7:36–46, 2018.

Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature methods*, 17(2):147–154, 2020.

Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J. L., Saunders, L., Trapnell, C., and Kannan, S. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell systems*, 10(3):265–274, 2020.

Rahimzamani, A. and Kannan, S. Network inference using directed information: The deterministic limit. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 156–163. IEEE, 2016.

Stumpf, M. P. Inferring better gene regulation networks from single-cell data. *Current Opinion in Systems Biology*, 27:100342, 2021.

Teschendorff, A. E. and Feinberg, A. P. Statistical mechanics meets single-cell biology. *Nature Reviews Genetics*, 22(7):459–476, 2021.

Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H., and Theis, F. J. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development*, 146(12):dev170506, 2019.

Wang, X., Choi, D., and Roeder, K. Constructing local cell-specific networks from single-cell data. *Proceedings of the National Academy of Sciences*, 118(51), 2021.

Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M., and Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.

Zhang, S., Afanassiev, A., Greenstreet, L., Matsumoto, T., and Schiebinger, G. Optimal transport analysis reveals trajectories in steady-state systems. *PLoS computational biology*, 17(12):e1009466, 2021.

Zhang, Z., Han, J., Song, L., and Zhang, X. Inferring cell-specific gene regulatory networks from single cell gene expression data. *bioRxiv*, 2022.

A. Appendix

A.1. Constructing a backward operator

It is straightforward to write:

$$Q_{ij} = \mathbf{P}[X_{-t} = x_j | X_0 = x_i] = \frac{\mathbf{P}[X_0 = x_i | X_{-t} = x_j]\mathbf{P}[X_{-t} = x_j]}{\mathbf{P}[X_0 = x_i]} = \frac{(P^t)_{ji}\mathbf{P}[X_{-t} = x_j]}{\mathbf{P}[X_0 = x_i]}$$

If P encoded a reversible Markov chain and we chose $\mathbf{P}[X_{-t} = \cdot]$ to be the stationary distribution, this would give us time-reversal at equilibrium. Since we are interested in the behaviour of the process away from equilibrium, we must prescribe $\mathbf{P}[X_{-t} = \cdot]$ away from equilibrium. In practice, we find that taking $\mathbf{P}[X_{-t} = \cdot] = \mathrm{Unif}$ works well – this is equivalent to simply taking the transpose of P and rescaling, as done in (Lange et al., 2022).

We refer the reader to the analysis presented in Section 4.3 of (Li et al., 2020) for a more detailed discussion on the transpose as a backward operator and its interpretation in the continuous limit.

A.2. Smooth and sparse optimization

The problem (4) is convex and a straightforward numerical algorithm for its solution can be derived using the alternating direction method of multipliers (ADMM) (Boyd et al., 2011). With auxiliary variables W, Z with the same dimensions as X, the corresponding ADMM scheme is

$$X^{(k+1)} = (\alpha + \lambda_1 L + \rho I)^{-1}(\alpha\hat{G} + \rho(Z^{(k)} - W^{(k)})) \tag{5}$$

$$Z^{(k+1)} = \arg\min_Z \lambda_2\|Z\|_1 + \frac{\rho}{2}\|X^{(k+1)} - Z^{(k)} + W^{(k)}\|_2^2 = \mathrm{prox}_{\rho^{-1}\lambda_2\|\cdot\|_1}(X^{(k+1)} + W^{(k)}) \tag{6}$$

$$W^{(k+1)} = W^{(k)} + X^{(k+1)} - Z^{(k+1)} \tag{7}$$

where $\rho > 0$ is the ADMM relaxation parameter (we take this to be 0.05), $\alpha = \mathrm{diag}(w)$ and

$$\mathrm{prox}_{\lambda\|\cdot\|_1}(x) = \mathrm{sgn}(x)(|x| - \lambda)_+,$$

interpreted elementwise.

A.3. Parameter choices for transition kernel reconstruction

Velocity kernel As described in the main text, the velocity kernel was constructed from evaluations of the drift term of the chemical Langevin SDE using the `cellrank.tl.kernels.DotProductScheme` class within CellRank (Lange et al., 2022), with $\sigma = 0.5$ and $k = 25$. A reversed-velocity kernel was constructed by swapping the roles of P (the forward kernel) and Q (the reverse kernel).

Optimal transport kernel Cells in the top 5% of pseudotime in each branch were assigned a negative flux rate. R_i so that each branch had a net negative flux of -12.5. Remaining cells were assigned a positive flux rate to satisfy the zero net flux requirement, $\sum_i R_i = 0$. StationaryOT (Zhang et al., 2021) with quadratically regularized optimal transport was applied with $\Delta t = 1.0, \varepsilon = 0.05$. The cost matrix was taken to be the matrix of pairwise graph distances constructed from a k-NN graph of the example dataset with $k = 25$.

Population balance analysis (PBA) kernel PBA (Weinreb et al., 2018) was applied with the same flux rates R_i as used in the optimal transport kernel, with neighbours $k = 25$ and diffusivity $D = 5.0$.

Diffusion pseudotime kernel The diffusion pseudotime (DPT) kernel was constructed using the `cellrank.tl.kernels.PseudotimeKernel` class within CellRank.

Undirected kernel We chose $P = Q = \Pi$, where Π is the neighbourhood kernel constructed using quadratically regularized optimal transport (Matsumoto et al., 2022).