

1 **I. Connection of the projection to local linear embedding (LLE)**

2 In this work, we use the neighborhood of a data point to approximate the corresponding tangent
3 space of the data manifold. Such approximation has been used in other applications such as the
4 local linear embedding (LLE)⁵. Here we provide a derivation of representing velocity vectors
5 based on LLE.

6 Consider a data point \mathbf{x}_i and its neighbors. Applying LLE one has $\sum_{j \neq i} w_{ij}(x_j - x_i) =$
7 $\sum_{j \neq i} w_{ij}\delta_{ij} = 0$, with $\sum_j w_{ij} = 1$. The LLE algorithm gives $w_{ij} = \frac{\sum_k c_{jk}^{-1}}{\sum_{lm} c_{lm}^{-1}}$, with $c_{ij} = (x_i - x_j) \cdot$
8 $(x_i - x_k)$. Now consider after $\delta t \rightarrow 0$ \mathbf{x}_i moves to $\mathbf{x}'_i = \mathbf{x}_i + \mathbf{v}_{pi}\delta t$, which should be also within
9 the local linear embedding space. Again with LLE one has $\sum_{j \neq i} w'_{ij}(x_j - x'_i) = 0$, with
10 $\sum_j w'_{ij} = 1$. Then one has $\mathbf{v}_{pi} = \sum_{j \neq i} \frac{(w'_{ij} - w_{ij})}{\delta t} \delta_{ij} = \sum_{j \neq i} \phi_{ij} \delta_{ij}$, with a form similar to that in eq.
11 1 of the main text.

12 The above derivation suggests an iterative trajectory propagation algorithm for obtaining the
13 velocities in the new representation:

- 14 1) Use a trial \mathbf{v}_{pi} to propagate the state and obtain $(\mathbf{x}_i, \mathbf{x}'_i)$. One can propagate backward and
15 forward by $\delta t/2$ and use central difference to estimate the velocity vectors.
- 16 2) Apply LLE to get w_{ij} and w'_{ij} . I
- 17 3) Obtain a new \mathbf{v}_{pi}
- 18 4) Go back to step 1 until the result converges below a threshold.

19 We numerically tested the above algorithm, the one with the loss function given in eq. 2 of the
20 main text with the parameter $b = 0$ and $b \neq 0$. All the three algorithms worked well on synthetic
21 data. The first two actually outperformed the third one (used in the main text). However, the first

22 two sometime showed numerical instability once applied to real single cell data. Therefore, in
23 subsequent applications, we used the general loss function form in eq. 2, using the direction
24 information from the second term (with $b \neq 0$) to further regularize the projection.

25 **II. Numerical issues of using eq. 1 for tangent space projection**

26 The relation in eq. 1 provides an algorithm for projecting a measured \mathbf{v} onto \mathcal{M} by minimizing the
27 following loss function,

28
$$\mathcal{L}(\boldsymbol{\phi}_i) = \|\mathbf{v}_i - \mathbf{v}_{\parallel i}\|^2 + \lambda \|\boldsymbol{\phi}_i\|^2,$$

29 which unfortunately is numerically unstable. The redundancy of the basis vectors leads to
30 coefficients that are not uniquely determined, and failure of the projection. To see the latter, notice
31 that in real data the subspace formed by the displacement vectors only approximates the tangent
32 space $T\mathcal{M}$ locally, and it likely contains small components in the orthogonal space. Then the
33 projection procedure tries to express both $\mathbf{v}_{\parallel}(\mathbf{x})$ and all or part of the remaining $(\mathbf{v}_i - \mathbf{v}_{\parallel})$ as a linear
34 combination of $\boldsymbol{\delta}_{ij}$ with some $|\phi_{ij}| \gg 0$, which can only be counterbalanced by a very large value
35 of λ . The latter would impose large weight on the regularization. While further data preprocessing
36 may be developed in the future to alleviate the problem, in this work we provided a strategy of
37 adding an additional cosine kernel term for the direction information.

38 We also experimented with the idea of removing the redundancy through dimension reduction
39 with PCA. The numerical results were not satisfactory, since it converged slowly with the sampling
40 size, which becomes impractical.

42 **III. Reformulation of Cosine kernel used in the literature in the context of tangent**
43 **space projection**

44 In the original RNA velocity study, a Cosine kernel has been proposed to transform the RNA
45 velocity vectors between different representations⁶. We re-casted the cosine kernel in the context
46 of the tangent space projection with a form similar to that of eq. 1 in the main text:

$$\mathbf{v}_{\parallel}^{\text{corr}}(\mathbf{x}_i) = \sum_{j \in \mathcal{N}_i} \phi_{ij}^{\text{corr}} \boldsymbol{\delta}_{ij},$$

47 where $\phi_{ij}^{\text{corr}} = P_{ij} - \frac{1}{k}$, and $P_{ij} = \frac{\exp[\cos(\mathbf{v}_i, \boldsymbol{\delta}_{ij})/\sigma]}{\sum_m \exp[\cos(\mathbf{v}_i, \boldsymbol{\delta}_{im})/\sigma]}$ defined in the form of softmax functions,
48 with $\cos(\cdot, \cdot)$ denoting the cosine similarity between two input vectors, σ an arbitrary bandwidth
49 parameter, and k the number of neighbors for each cell. Here P_{ij} gives a heuristic transition
50 probability from cell i to j . The term $(-1/k)$, called the “density correction”, is designed to correct
51 the potential sampling bias where the embedded velocity vectors tend to point towards the
52 direction of regions with high cell density. Li et al. showed that mathematically the cosine kernel
53 asymptotically converges to the correct direction of a velocity vector⁷. However, the correlation
54 kernel loses information about the magnitude of the velocity vectors \mathbf{v}_i (i.e., the *speed*), due to the
55 normalization in the correlation functions. Intuitively, the correlation kernel is qualitatively guided
56 by the physical intuition that a cell has a high tendency to move along the direction of its velocity
57 vector. Here, we used the direction information of ϕ_{ij}^{corr} to help on constraining $\boldsymbol{\Phi}_i$.

58 **IV. Further discussions on applications of GraphVelo to manifolds formed by multi-**
59 **modal single cell data.**

60 GraphVelo is based on the following mathematical assumption: the manifolds of a given system
61 embedded in two different spaces are homeomorphic so one can establish a one-to-one mapping

62 between the two. Here we use one simple example of scRNAseq/scATACseq multiomic data to
63 illustrate that one can still apply GraphVelo if this assumption is violated. Consider that a cellular
64 system switches between two discrete epigenetic states, and for each epigenetic state there is a
65 corresponding (quasi)continuous transcriptomic manifold. Label the two disjoint manifolds as
66 $\mathcal{M}_1(\mathbf{x})$ and $\mathcal{M}_2(\mathbf{x})$, where \mathbf{x} represents the transcriptomic state. Consider two cells having the same
67 \mathbf{x} but different chromatin state, c_1 on \mathcal{M}_1 and c_2 on \mathcal{M}_2 . Multiomic data allows distinction between
68 two cells. That is, the neighborhood of c_1 is composed of cells on \mathcal{M}_1 , and the neighborhood of
69 c_2 is composed of cells on \mathcal{M}_2 . Therefore, GraphVelo analyses treat cells on \mathcal{M}_1 and \mathcal{M}_2
70 separately.

71 **V. Feature selection for robustly estimated velocity genes.**

72 In datasets featuring non-differentiating cell types organized in a hierarchical manner, application
73 of RNA velocity may show different kinetic regimes. Analyzing cell fate transition using
74 unreliable velocity genes can lead to emergence of spurious cell state transitions. Different
75 packages subset the well-estimated kinetic genes for further analyses in different scenarios. In the
76 implementation of scVelo, one first regresses out the uncertain genes are based on the r-squared
77 coefficient in the steady-state model and further selects trustful velocity genes according to the
78 log-likelihood in the dynamical mode. PhyloVelo relies on monotonically expressed genes to infer
79 transcriptomic velocity, which is a more stringent requirement. However, the monotonic
80 assumptions are counter intuitive and oversimplify the gene expression kinetics along
81 differentiation or disease trajectories due to the non-sequential nature of these cascades. Dynamo
82 provides a biological prior-guided method and filters out the unreliable genes based on the
83 knowledge of lineage information.

85 1. Gillespie, D.T. The chemical Langevin Equation. *J. Chem. Phys.* **113**, 297-306 (2000).

86 2. Zwanzig, R. Ensemble method in the theory of irreversibility. *J. Chem. Phys.* **33**, 1338-1341
87 (1960).

88 3. Mori, H. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* **33**, 423-455
89 (1965).

90 4. Xing, J. & Kim, K.S. Application of the projection operator formalism to non-Hamiltonian
91 dynamics. *Journal of Chemical Physics* **134**(2011).

92 5. Roweis, S.T. & Saul, L.K. Nonlinear Dimensionality Reduction by Locally Linear Embedding.
93 *Science* **290**, 2323-2326 (2000).

94 6. La Manno, G. *et al.* RNA velocity of single cells. *Nature* (2018).

95 7. Li, T., Shi, J., Wu, Y. & Zhou, P. On the Mathematics of RNA Velocity I: Theoretical Analysis.
96 *CSIAM Transactions on Applied Mathematics* **2**, 1-55 (2021).

97