

MATHEMATICALLY PROVED

Considering the technical variability or measurement error in the transcriptomic data [1], [2], [3], it is important to examine the robustness of the method with respect to the perturbation in pseudotime progression. In this part, the robustness of model to the measured variables in the data is illustrated by mathematical proof.

Theorem 1. Assume there are two trajectories of pseudotime progression $s(r)$ and $\tilde{s}(r)$ with the same root $r \in I = [0, 1]$. Define $\|\tilde{s} - s\|_{L^2} = (\int_I |\tilde{s} - s|^2 dr)^{1/2}$. If $(X_i(s), U_l(s), a_{ij}, b_{il}, c_{lk})$ and $(X_i(\tilde{s}), U_l(\tilde{s}), \tilde{a}_{ij}, \tilde{b}_{il}, \tilde{c}_{lk})$ both satisfy the equations of progression-dependent dynamical system, i.e.,

$$\frac{dX_i(s)}{ds} = \sum_{j \neq i} a_{ij} X_j(s) \cdot X_j(s) + \sum_{l=1}^M b_{il} X_i(s) \cdot U_l(s) - d_i X_i(s), \quad i = 1, 2, \dots, N, \quad (1)$$

$$\frac{dU_l(s)}{ds} = \sum_{k \neq l} c_{lk} U_l(s) \cdot U_k(s) - d'_l U_l(s), \quad l = 1, 2, \dots, M, \quad (2)$$

$$\frac{dX_i(\tilde{s})}{d\tilde{s}} = \sum_{j \neq i} \tilde{a}_{ij} X_j(\tilde{s}) \cdot X_j(\tilde{s}) + \sum_{l=1}^M \tilde{b}_{il} X_i(\tilde{s}) \cdot U_l(\tilde{s}) - \tilde{d}_i X_i(\tilde{s}), \quad i = 1, 2, \dots, N, \quad (3)$$

$$\frac{dU_l(\tilde{s})}{d\tilde{s}} = \sum_{k \neq l} \tilde{c}_{lk} U_l(\tilde{s}) \cdot U_k(\tilde{s}) - \tilde{d}'_l U_l(\tilde{s}), \quad l = 1, 2, \dots, M, \quad (4)$$

then we have

$$\lim_{\|\tilde{s}-s\|_{L^2} \rightarrow 0} \left(\sum_{j=1}^N (\tilde{a}_{ij} - a_{ij})^2 + \sum_{l=1}^M (\tilde{b}_{il} - b_{il})^2 \right) = 0, \quad (5)$$

and

$$\lim_{\|\tilde{s}-s\|_{L^2} \rightarrow 0} \left(\sum_{k=1}^M (\tilde{c}_{lk} - c_{lk})^2 \right) = 0. \quad (6)$$

Proof: We first transform the (1) into the following form,

$$\frac{d(\log X_i(s))}{ds} = \sum_{j \neq i} a_{ij} X_j(s) + \sum_{l=1}^M b_{il} U_l(s) - d_i, \quad (7)$$

Denoting $a_{ii} = 0$ and integrating the above equation leads to

$$\log X_i(s) = \sum_{j=1}^N a_{ij} \int_0^s X_j(\tau) d\tau + \sum_{l=1}^M b_{il} \int_0^s U_l(\tau) d\tau - d_i s + \log X_i(0). \quad (8)$$

Denote $Y_i(s) = \log X_i(s)$, $Z_j(s) = \int_0^s X_j(\tau) d\tau$ and $H_l(s) = \int_0^s U_l(\tau) d\tau$. The above equation can be written as

$$\begin{aligned}
Y_i(s) &= \sum_{j=1}^N a_{ij} Z_j(s) + \sum_{l=1}^M b_{il} H_l(s) - d_i s + Y_i(0) \\
&= \sum_{j=1}^N (a_{ij} - \tilde{a}_{ij}) Z_j(s) + \sum_{j=1}^N \tilde{a}_{ij} Z_j(s) + \sum_{l=1}^M (b_{il} - \tilde{b}_{il}) H_l(s) + \sum_{l=1}^M \tilde{b}_{il} H_l(s) - d_i s + Y_i(0) \\
&= \sum_{j=1}^N (a_{ij} - \tilde{a}_{ij}) Z_j(s) + \sum_{l=1}^M (b_{il} - \tilde{b}_{il}) H_l(s) + \sum_{j=1}^N \tilde{a}_{ij} (Z_j(s) - Z_j(\tilde{s})) + \\
&\quad \sum_{l=1}^M \tilde{b}_{il} (H_l(s) - H_l(\tilde{s})) + \left[\sum_{j=1}^N \tilde{a}_{ij} Z_j(\tilde{s}) + \sum_{l=1}^M \tilde{b}_{il} H_l(\tilde{s}) - \tilde{d}_i \tilde{s} + Y_i(0) \right] + \tilde{d}_i \tilde{s} - d_i s \\
&= \sum_{j=1}^N (a_{ij} - \tilde{a}_{ij}) Z_j(s) + \sum_{l=1}^M (b_{il} - \tilde{b}_{il}) H_l(s) + Y_i(\tilde{s}) + \sum_{j=1}^N \tilde{a}_{ij} (Z_j(s) - Z_j(\tilde{s})) + \\
&\quad \sum_{l=1}^M \tilde{b}_{il} (H_l(s) - H_l(\tilde{s})) - \tilde{d}_i (s - \tilde{s}) - (d_i - \tilde{d}_i) s.
\end{aligned} \tag{9}$$

Therefore, we have

$$\begin{aligned}
&\sum_{j=1}^N (a_{ij} - \tilde{a}_{ij}) Z_j(s) + \sum_{l=1}^M (b_{il} - \tilde{b}_{il}) H_l(s) - (d_i - \tilde{d}_i) s \\
&= Y_i(s) - Y_i(\tilde{s}) - \sum_{j=1}^N \tilde{a}_{ij} (Z_j(s) - Z_j(\tilde{s})) - \sum_{l=1}^M \tilde{b}_{il} (H_l(s) - H_l(\tilde{s})) + \tilde{d}_i (s - \tilde{s}).
\end{aligned} \tag{10}$$

For $i = 1, 2, \dots, N$ and $j = 1, \dots, N + M$, define

$$w_{ij} = \begin{cases} a_{ij} - \tilde{a}_{ij}, & i \neq j, \quad j = 1, \dots, N \\ -(d_i - \tilde{d}_i), & i = j, \quad j = 1, \dots, N \\ b_{i,j-N} - \tilde{b}_{i,j-N}, & j = N + 1, \dots, N + M \end{cases}$$

$$\text{and redefine } Z_j(s) = \begin{cases} \int_0^s X_j(\tau) d\tau, & i \neq j, \quad j = 1, \dots, N \\ s, & i = j, \quad j = 1, \dots, N \\ \int_0^s U_{j-N}(\tau) d\tau, & j = N + 1, \dots, N + M \end{cases}$$

Denote $W_i = (w_{i1}, w_{i2}, \dots, w_{i,N+M})^T$, $Z = (Z_1(s), Z_2(s), \dots, Z_{N+M}(s))$ and $V_i = Y_i(s) - Y_i(\tilde{s}) - \sum_{j=1}^N \tilde{a}_{ij} (Z_j(s) - Z_j(\tilde{s})) - \sum_{l=1}^M \tilde{b}_{il} (H_l(s) - H_l(\tilde{s})) + \tilde{d}_i (s - \tilde{s})$. Then (10) can be written as

$$ZW_i = V_i. \quad (11)$$

Take $m + 1$ points on $s(r)$ and $\tilde{s}(r)$, respectively. Here, $s_i = s(r_i)$ and $\tilde{s}_i = \tilde{s}(r_i)$, where $r_i = \frac{i}{m}$, $i = 0, 1, \dots, m$. Let $U_i = (V_i(r_0), V_i(r_1), \dots, V_i(r_{m-1}))^T$ and

$$T = \begin{bmatrix} Z_1(s_0) & Z_2(s_0) & \cdots & Z_{N+M}(s_0) \\ Z_1(s_1) & Z_2(s_1) & \cdots & Z_{N+M}(s_1) \\ \vdots & \vdots & \ddots & \vdots \\ Z_1(s_{m-1}) & Z_2(s_{m-1}) & \cdots & Z_{N+M}(s_{m-1}) \end{bmatrix}_{m \times (N+M)}. \quad (12)$$

Then we have

$$TW_i = U_i. \quad (13)$$

Therefore, we have

$$\hat{W}_i = \arg \min_{W_i} \|U_i - TW_i\|^2 = \arg \min_{W_i} (U_i - TW_i)^T (U_i - TW_i). \quad (14)$$

By matrix derivation, we get

(a) When $\text{rank}(T) = N + M$,

$$\hat{W}_i = (T^T T)^{-1} T^T U_i. \quad (15)$$

(b) When $\text{rank}(T) \neq N + M$, there exists $\lambda > 0$ such that

$$\hat{W}_i = (T^T T + \lambda I_{N+M})^{-1} T^T U_i. \quad (16)$$

From the definition of W_i , (15) and (16), we know that

$$\sum_{j=1}^N (a_{ij} - \tilde{a}_{ij})^2 + \sum_{l=1}^M (b_{il} - \tilde{b}_{il})^2 + (d_i - \tilde{d}_i)^2 \leq \alpha_1 \cdot \|U_i\|_{L^2}^2, \quad (17)$$

where α_1 is a positive constant. According to the definition of U_i , we have

$$\|U_i\|_{L^2} \leq \alpha_2 \|V_i\|_{L^2}, \quad (18)$$

where α_2 is a positive constant and $\|V_i\|_{L^2} \leq \|Y_i(s) - Y_i(\tilde{s})\|_{L^2} + \|\sum_{j=1}^N \tilde{a}_{ij} (Z_j(s) - Z_j(\tilde{s}))\|_{L^2} + \|\sum_{l=1}^M \tilde{b}_{il} (H_l(s) - H_l(\tilde{s}))\|_{L^2} + \|\tilde{d}_i\|_{L^2} \|s - \tilde{s}\|_{L^2}$. Now we only need to prove that the first three terms in the right hand of the above inequality tends to 0 as $\|s - \tilde{s}\| \rightarrow 0$.

- (i) We first prove $\lim_{\|s-\tilde{s}\| \rightarrow 0} \|Y_i(s) - Y_i(\tilde{s})\|_{L^2} = 0$. Since $Y_i(s) = \log X_i(s)$, we only need to prove $\lim_{\|s-\tilde{s}\| \rightarrow 0} \|X_i(s) - X_i(\tilde{s})\|_{L^2} = 0$, which is valid according to the continuity of $X_i(s)$ with respect to s .
- (ii) Then we prove $\lim_{\|s-\tilde{s}\| \rightarrow 0} \left\| \sum_{j=1}^N \tilde{a}_{ij}(Z_j(s) - Z_j(\tilde{s})) \right\|_{L^2} = 0$.

$$\begin{aligned}
& \sum_{j=1}^N \tilde{a}_{ij}(Z_j(s) - Z_j(\tilde{s})) \\
&= - \sum_{j=1}^N \tilde{a}_{ij} \left(\int_0^{\tilde{s}} X_j(\tau) d\tau - \int_0^s X_j(\tau) d\tau \right) \\
&= - \sum_{j=1}^N \tilde{a}_{ij} \left[\lim_{m \rightarrow \infty} \sum_{l=0}^{m-1} \left(\int_{\tilde{s}_l}^{\tilde{s}_{l+1}} X_j(\tau) d\tau - \int_{s_l}^{s_{l+1}} X_j(\tau) d\tau \right) \right].
\end{aligned} \tag{19}$$

For any m , since $\|s - \tilde{s}\| \rightarrow 0$, we have $\tilde{s}_l \rightarrow s_l$, $l = 0, 1, \dots, m$ and thus

$$\begin{aligned}
& \lim_{\|s-\tilde{s}\| \rightarrow 0} \sum_{l=0}^{m-1} \left(\int_{\tilde{s}_l}^{\tilde{s}_{l+1}} X_j(\tau) d\tau - \int_{s_l}^{s_{l+1}} X_j(\tau) d\tau \right) \\
&= \sum_{l=0}^{m-1} \lim_{\|s-\tilde{s}\| \rightarrow 0} \left(\int_{\tilde{s}_l}^{\tilde{s}_{l+1}} X_j(\tau) d\tau - \int_{s_l}^{s_{l+1}} X_j(\tau) d\tau \right).
\end{aligned} \tag{20}$$

That is, the above limit is uniformly convergent with respect to m . Therefore

$$\begin{aligned}
& \lim_{\|s-\tilde{s}\| \rightarrow 0} \left(\int_0^{\tilde{s}} X_j(\tau) d\tau - \int_0^s X_j(\tau) d\tau \right) \\
&= \lim_{\|s-\tilde{s}\| \rightarrow 0} \lim_{m \rightarrow \infty} \sum_{l=0}^{m-1} \left(\int_{\tilde{s}_l}^{\tilde{s}_{l+1}} X_j(\tau) d\tau - \int_{s_l}^{s_{l+1}} X_j(\tau) d\tau \right) \\
&= \lim_{m \rightarrow \infty} \sum_{l=0}^{m-1} \lim_{\|s-\tilde{s}\| \rightarrow 0} \left(\int_{\tilde{s}_l}^{\tilde{s}_{l+1}} X_j(\tau) d\tau - \int_{s_l}^{s_{l+1}} X_j(\tau) d\tau \right) \\
&= 0.
\end{aligned} \tag{21}$$

From the above (19), (20) and (21), we can obtain that

$$\lim_{\|s-\tilde{s}\| \rightarrow 0} \left\| \sum_{j=1}^N \tilde{a}_{ij}(Z_j(s) - Z_j(\tilde{s})) \right\|_{L^2} = 0.$$

- (iii) Similarly, we have $\lim_{\|s-\tilde{s}\| \rightarrow 0} \left\| \sum_{l=1}^M \tilde{b}_{il}(H_l(s) - H_l(\tilde{s})) \right\|_{L^2} = 0$.

Therefore, (5) holds. Similarly, we can prove that (6) holds. ■

Based on the spectral graph theory [4], [5], the above manifold distance is noise-resistant, so the variation in the progression trajectory, i.e., $\tilde{s} - s$, should be small considering the moderate perturbations. As a result, Theorem 1 means that the corresponding estimates of $[a_{ij}]_{N \times N}$, $[b_{il}]_{N \times M}$

and $[c_{lk}]_{M \times M}$ should vary minimally. Therefore, the above theorem theoretically ensures the consistency and robustness of the regulatory coefficient estimation. Moreover, the Bayesian Lasso method used in this paper can further guarantee a robust implementation of network inference.

A corollary of the above theorem is that the mappings $s \mapsto [a_{ij}(s)]_{N \times N}$, $s \mapsto [b_{il}(s)]_{N \times M}$ and $s \mapsto [c_{lk}(s)]_{M \times M}$ are continuous under some appropriate metric. More specifically, for two trajectories s and $\tilde{s}$, if the difference between the two inferred regulatory coefficients $[a_{ij}(s)]_{N \times N}$, $[b_{il}(s)]_{N \times M}$, $[c_{lk}(s)]_{M \times M}$ and $[a_{ij}(\tilde{s})]_{N \times N}$, $[b_{il}(\tilde{s})]_{N \times M}$, $[c_{lk}(\tilde{s})]_{M \times M}$ is notably greater than 0, then the difference between s and $\tilde{s}$ should not be arbitrarily small. This means that if the inferred regulatory networks for two progressions are largely different, then the two progressions should have distinct trajectories, resulting in distinct different clinical outcomes. Therefore, Theorem 1 also shows that our proposed model may be used to predict or control cancer progression.

REFERENCES

- [1] Ji,H., and Davis,R.W. (2006). Data quality in genomics and microarrays, *Nature Biotechnology*, **24(9)**, 1112–1113.
- [2] Nie,L. *et al.* (2007). Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications, *Critical Reviews in Biotechnology*, **27(2)**, 63–75.
- [3] Su,Z. *et al.* (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium, *Nature Biotechnology*, **32(9)**, 903–914.
- [4] Chung,F.R., and Graham,F.C. (1997). *Spectral graph theory*, American Mathematical Soc.
- [5] Coifman,R.R., and Lafon,S. (2006). Diffusion maps, *Applied and Computational Harmonic Analysis*, **21(1)**, 5–30.