Network Reconstruction via High-Dimensional ODEs

Ali Shojaie
Department of Biostatistics
University of Washington
`faculty.washington.edu/ashojaie`

SIAM Annual Meeting – 2017
Joint work with Shizhe Chen & Daniela Witten

Nonlinear Dynamics in Gene Regulatory Networks

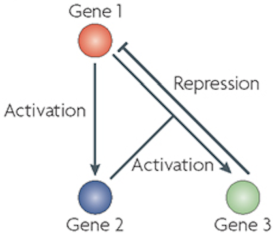
a

$$\frac{d(gene_1)}{dt} = k_{1,s} \cdot \frac{1}{1 + k_{1,3} \cdot gene_3} - k_{1,d} \cdot gene_1$$

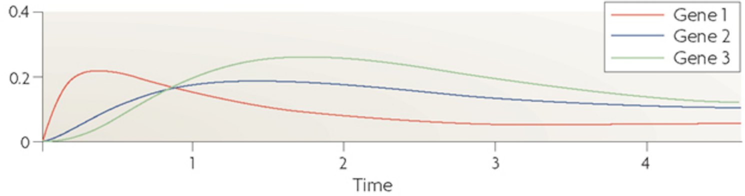
$$\frac{d(gene_2)}{dt} = k_{2,s} \cdot \frac{k_{2,1} \cdot gene_1}{1 + k_{2,1} \cdot gene_1} - k_{2,d} \cdot gene_2$$

$$\frac{d(gene_3)}{dt} = k_{3,s} \cdot \frac{k_{3,1} \cdot gene_1 \cdot k_{3,2} \cdot gene_2}{(1 + k_{3,1} \cdot gene_1) \cdot (1 + k_{3,2} \cdot gene_2)} - k_{3,d} \cdot gene_3$$

b



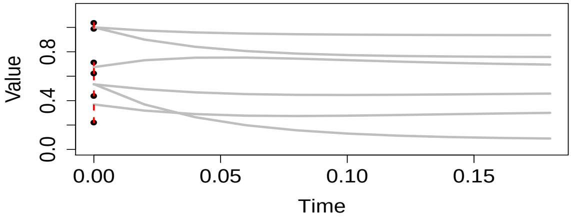
c



Karlebach & Shamir (2008), Nature Reviews: Molecular Cell Biology

Time-course gene expression data

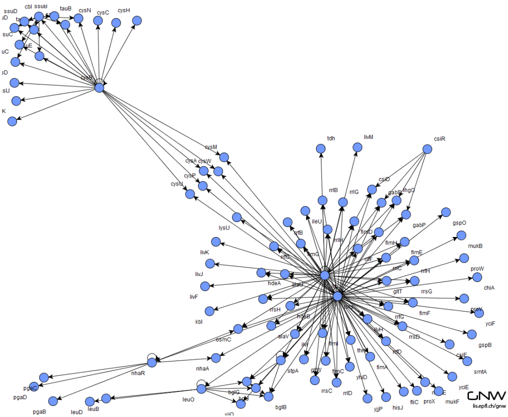
- p genes
- Expressions $X(\cdot)$ measured at n discrete time points $t_1, \ldots, t_n$



Measurments

	0	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.1
rseB	0.22									
cpxR	0.99									
motB	0.71									
mdtA	0.62									
ompC	0.44									
ihfB	1.04									

E. coli Gene Regulatory Network[1]



In this graph: $X_1 \longrightarrow X_2 \iff X_1$ regulates X_2

[1] Subnetwork of *E-coli* regulatory network (GeneNetWeaver, Schaffter et al, 2011)

Gene Regulatory Network as a System of ODEs

Gene Regulatory Network as a System of ODEs

The change in expression of one gene is "regulated" by the expressions of others at the same time point

$$X_j'(t) = f_j(X(t), \theta_j)$$

Gene Regulatory Network as a System of ODEs

The change in expression of one gene is “regulated” by the expressions of others at the same time point

$$X_j'(t) = f_j(X(t), \theta_j)$$

Example: Linear ODEs, $X'(t) = \Theta X(t) \quad (\Rightarrow X(t) = \exp(\Theta t)X(0))$

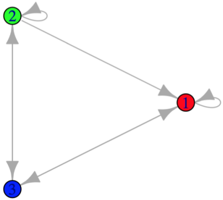
Gene Regulatory Network as a System of ODEs

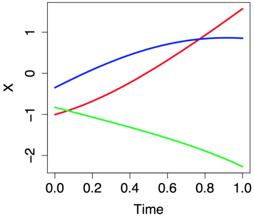
The change in expression of one gene is “regulated” by the expressions of others at the same time point

$$X_j'(t) = f_j(X(t), \theta_j)$$

Example: Linear ODEs, $X'(t) = \Theta X(t)$ $\quad (\Rightarrow X(t) = \exp(\Theta t)X(0))$

$$\Theta = \begin{pmatrix} -0.59 & -1.36 & 1.32 \\ 0.00 & 1.18 & 0.62 \\ -1.52 & -0.93 & 0.00 \end{pmatrix}$$





Estimation of ODEs from Noisy Observations

Estimation of ODEs from Noisy Observations

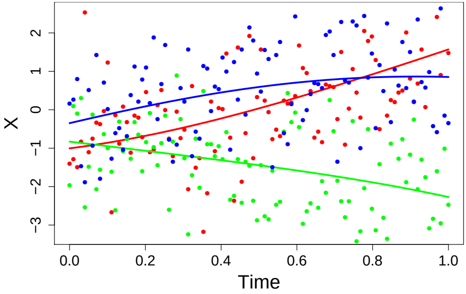
- Given $X(t)$ (and hence $X'(t)$), can find Θ ... either closed-form solutions or numerical methods

Estimation of ODEs from Noisy Observations

- Given $X(t)$ (and hence $X'(t)$), can find Θ ... either closed-form solutions or numerical methods
- We observe noisy measurements at discrete time points: $Y_i = X(t_i) + \varepsilon_i$

Estimation of ODEs from Noisy Observations

- Given $X(t)$ (and hence $X'(t)$), can find Θ ... either closed-form solutions or numerical methods
- We observe noisy measurements at discrete time points: $Y_i = X(t_i) + \varepsilon_i$



Existing Approaches: Parameter Estimation

Existing Approaches: Parameter Estimation

Suppose f is known but θ is unknown

$$X'(t) = f\big(X(t), \theta\big)$$

Existing Approaches: Parameter Estimation

Suppose f is known but θ is unknown

$$X'(t) = f\big(X(t), \theta\big)$$

<u>Gold Standard</u>:

- Find θ such that $X(\cdot;\theta)$ solves the ODE:

$$\hat{\theta} = \arg\min_\theta \sum_{i=1}^{n} \|Y_i - X(t_i;\theta)\|^2$$
$$\text{s.t. } X'(t;\theta) = f\left(X(t;\theta), \theta\right)$$

Existing Approaches: Parameter Estimation

Suppose f is known but θ is unknown

$$X'(t) = f\big(X(t), \theta\big)$$

Gold Standard:

- Find θ such that $X(\cdot;\theta)$ solves the ODE:

$$\hat{\theta} = \arg\min_\theta \sum_{i=1}^{n} \|Y_i - X(t_i;\theta)\|^2$$

$$\text{s.t. } X'(t;\theta) = f\,(X(t;\theta), \theta)$$

- Accurate but slow
  - requires numerical solution of ODE for every candidate θ
  - $\sqrt{n}$-consistent[2]

[2] e.g., Hall & Ma (2014)

Existing Approaches: Parameter Estimation

Suppose f is known but θ is unknown

$$X'(t) = f\big(X(t), \theta\big)$$

Existing Approaches: Parameter Estimation

Suppose f is known but θ is unknown

$$X'(t) = f\big(X(t), \theta\big)$$

Collocation Methods[3]:

- Two-stage estimation strategy:
  - estimate $\hat{X}(t)$ and $\hat{X}'(t)$ from data (e.g. kernel smoothing)
  - find θ that minimizes deviation from ODE:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \|\hat{X}'(t_i) - f\left(\hat{X}(t_i); \theta\right)\|_2^2$$

[3] Varah (1982)

Existing Approaches: Parameter Estimation

Suppose f is known but θ is unknown

$$X'(t) = f\big(X(t), \theta\big)$$

Collocation Methods[3]:

- Two-stage estimation strategy:
  - estimate $\hat{X}(t)$ and $\hat{X}'(t)$ from data (e.g. kernel smoothing)
  - find θ that minimizes deviation from ODE:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \|\hat{X}'(t_i) - f\left(\hat{X}(t_i); \theta\right)\|_2^2$$

- Fast but not exact
  - easy computation
  - not solving the ODE exactly!

[3]Varah (1982)

Gene Regulatory Network as a System of ODEs

Gene Regulatory Network as a System of ODEs

Challenges:

- Many genes and not many observations ($p \gg n$)
- Exact form of f not known!

$$X_j'(t) = f_j\left(X(t); \theta\right), \quad j = 1, \ldots, p$$

Gene Regulatory Network as a System of ODEs

Challenges:

- Many genes and not many observations ($p \gg n$)
- Exact form of f not known!

$$X_j'(t) = f_j\left(X(t);\theta\right), \quad j = 1, \ldots, p$$

Solution:

- Assume that f_j is additive

$$X_j'(t) = \sum_{k=1}^{p} f_{jk}(X_k(t);\theta), \quad j = 1, \ldots, p$$

$$Y_i = X(t_i) + \varepsilon_i, \ i = 1, \ldots, n$$

Gene Regulatory Network as a System of ODEs

Challenges:

- Many genes and not many observations ($p \gg n$)
- Exact form of f not known!

$$X_j'(t) = f_j\left(X(t);\theta\right), \quad j = 1,\ldots,p$$

Solution:

- Assume that f_j is additive

$$X_j'(t) = \sum_{k=1}^{p} f_{jk}(X_k(t);\theta), \quad j = 1,\ldots,p$$

$$Y_i = X(t_i) + \varepsilon_i,\ i = 1,\ldots,n$$

- $X_k \longrightarrow X_j$ iff $f_{jk} \neq 0$

Existing Approaches: Non-parametric Estimation

Suppose f_j and θ are both unknown: $X_j'(t) = f_j\big(X(t), \theta\big)$

Existing Approaches: Non-parametric Estimation

Suppose f_j and θ are both unknown: $X_j'(t) = f_j\big(X(t), \theta\big)$

Gradient Matching[4]:

- Estimate $\hat{X}_j(t)$ from Y_j (e.g., via nonparametric regression, etc)
- Calculate $\hat{X}_j'(t) \equiv \partial \hat{X}_j / \partial t$ (similar to collocation)

[4] Wu et al (2014), Henderson & Michailidis (2014)

Existing Approaches: Non-parametric Estimation

Suppose f_j and θ are both unknown: $X_j'(t) = f_j\big(X(t), \theta\big)$

Gradient Matching[4]:

- Estimate $\hat{X}_j(t)$ from Y_j (e.g., via nonparametric regression, etc)
- Calculate $\hat{X}_j'(t) \equiv \partial \hat{X}_j / \partial t$ (similar to collocation)
- For a truncated basis $\psi = (\psi_1, \ldots, \psi_M)^{\mathrm{T}}$,

$$f_{jk}(\cdot) = \psi(\cdot)^{\mathrm{T}} \theta_{jk} + \delta_{jk}(\cdot) \qquad \text{(allowing } M \to \infty \text{ with } n\text{)}$$

[4] Wu et al (2014), Henderson & Michailidis (2014)

Existing Approaches: Non-parametric Estimation

Suppose f_j and θ are both unknown: $X_j'(t) = f_j\big(X(t), \theta\big)$

Gradient Matching[4]:

- Estimate $\hat{X}_j(t)$ from Y_j (e.g., via nonparametric regression, etc)
- Calculate $\hat{X}_j'(t) \equiv \partial \hat{X}_j / \partial t$ (similar to collocation)
- For a truncated basis $\psi = (\psi_1, \ldots, \psi_M)^{\mathrm{T}}$,

$$f_{jk}(\cdot) = \psi(\cdot)^{\mathrm{T}} \theta_{jk} + \delta_{jk}(\cdot) \qquad \text{(allowing } M \to \infty \text{ with } n\text{)}$$

- For $j = 1, \ldots, p$, find $\hat{\theta}_j$ that minimizes

$$\int_0^1 \Big\{ \hat{X}_j'(t) - \underbrace{\theta_{j0} - \sum_{k=1}^{p} \psi\big(\hat{X}_k(t)\big)^{\mathrm{T}} \theta_{jk}}_{\hat{f}_j} \Big\}^2 dt + \lambda \sum_{k=1}^{p} \underbrace{\left[\int_0^1 \Big\{ \psi\big(\hat{X}_k(t)\big)^{\mathrm{T}} \theta_{jk} \Big\}^2 dt \right]^{1/2}}_{\text{group lasso penalty}}$$

A penalized regression problem!

[4] Wu et al (2014), Henderson & Michailidis (2014)

Key Observation – I

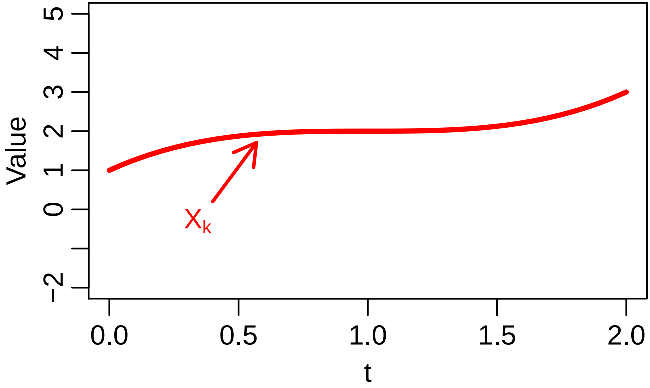
Key Observation – I

- Estimating the derivative is inefficient[5]!
- Hard to pick optimal bandwidth for estimating $d\hat{X}/dt$!

[5] More later...

Key Observation – I

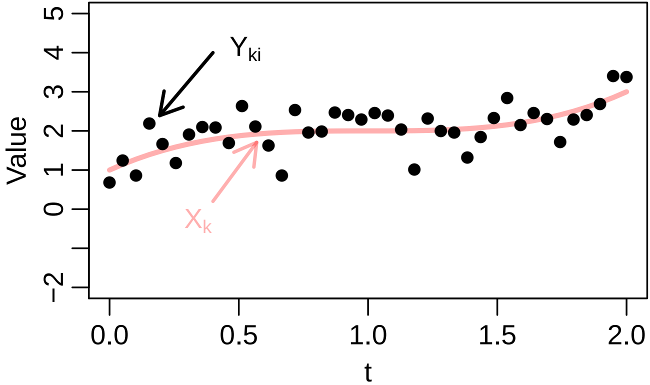
- Estimating the derivative is inefficient[5]!
- Hard to pick optimal bandwidth for estimating $d\hat{X}/dt$!



[5]More later...

Key Observation – I

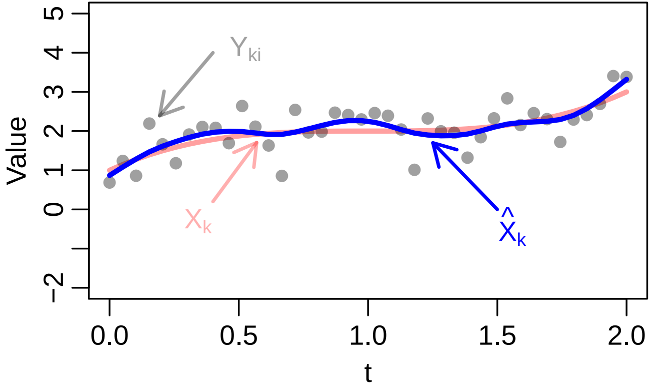
- Estimating the derivative is inefficient[5]!
- Hard to pick optimal bandwidth for estimating $d\hat{X}/dt$!



[5] More later...

Key Observation – I

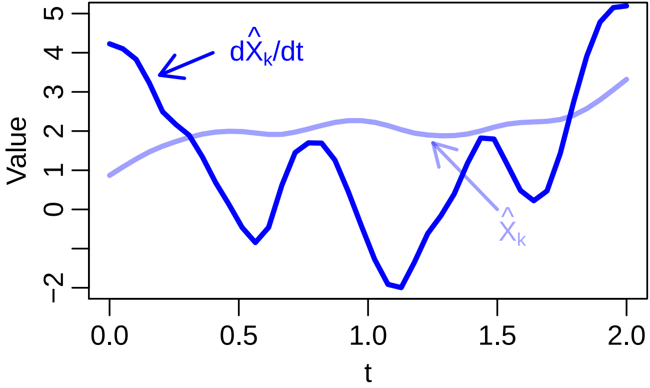
- Estimating the derivative is inefficient[5]!
- Hard to pick optimal bandwidth for estimating $d\hat{X}/dt$!



[5] More later...

Key Observation – I

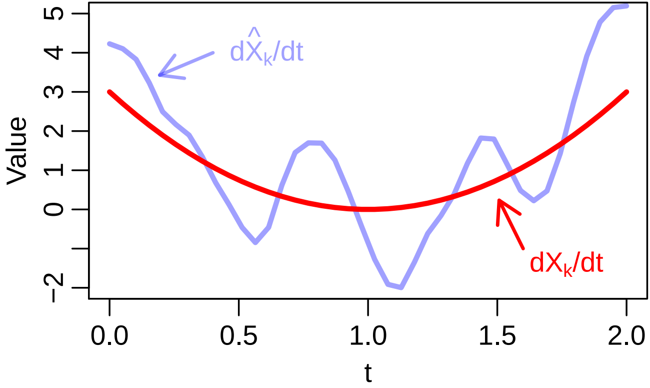
- Estimating the derivative is inefficient[5]!
- Hard to pick optimal bandwidth for estimating $d\hat{X}/dt$!



[5] More later...

Key Observation – I

- Estimating the derivative is inefficient[5]!
- Hard to pick optimal bandwidth for estimating $d\hat{X}/dt$!



[5] More later...

Key Observation – II

Key Observation – II

- Recall that

$$X_j'(t) \approx \theta_{j0} + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}} \theta_{jk}$$

Key Observation – II

- Recall that

$$X_j'(t) \approx \theta_{j0} + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}} \theta_{jk}$$

- Integrating both sides, we get

$$X_j(t) - X(0) \approx t\theta_{j0} + \sum_{k=1}^{p} \left(\int_0^t \psi\left(X_k(s)\right)^{\mathrm{T}} ds \right) \theta_{jk}$$

Key Observation – II

- Recall that

$$X_j'(t) \approx \theta_{j0} + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

- Integrating both sides, we get

$$X_j(t) - X(0) \approx t\theta_{j0} + \sum_{k=1}^{p} \left(\int_0^t \psi\big(X_k(s)\big)^{\mathrm{T}} ds \right) \theta_{jk}$$

- Let $\Psi_{ik} = \int_0^{t_i} \psi\big(X_k(s)\big)\, ds$,

$$Y_{ij} \approx X(0) + t_i\theta_{j0} + \sum_{k=1}^{p} \Psi_{ik}^{\mathrm{T}}\theta_{jk} + \varepsilon_{ij}$$

- Ψ_{ik} can be estimated as $\int_0^{t_i} \psi\big(\hat{X}_k(s)\big)\, ds$

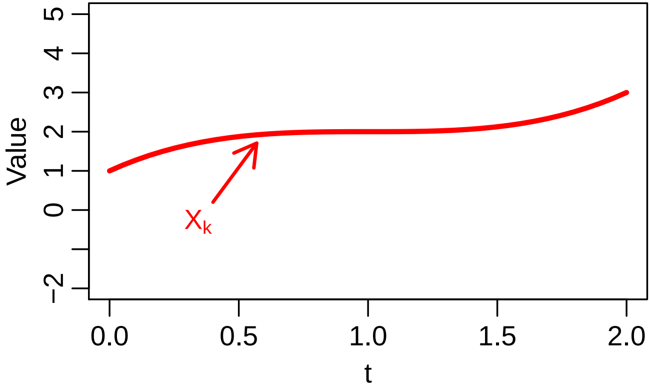
Key Observation – II

Key Observation – II

- Integral estimation is more precise!

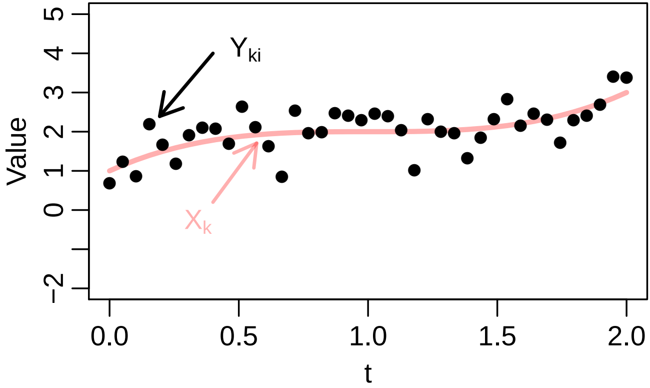
Key Observation – II

- Integral estimation is more precise!



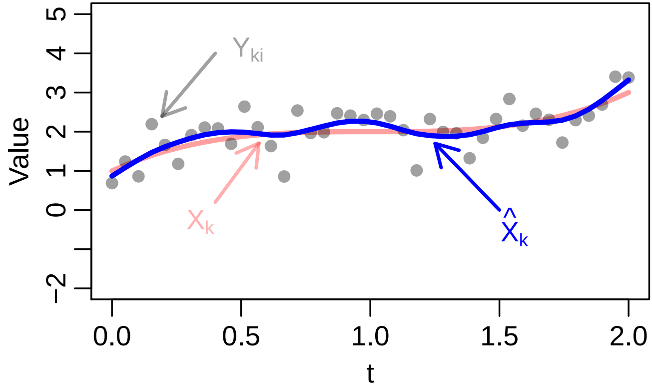
Key Observation – II

- Integral estimation is more precise!



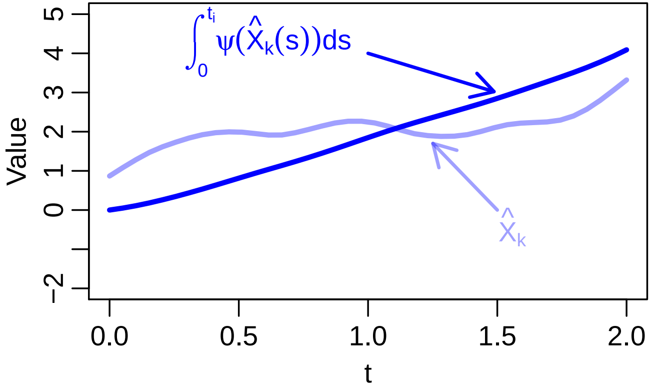
Key Observation – II

- Integral estimation is more precise!



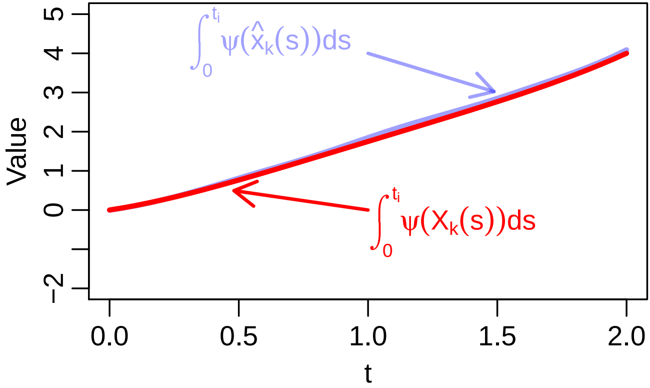
Key Observation – II

- Integral estimation is more precise!



Key Observation – II

- Integral estimation is more precise!



Our Proposal

Our Proposal

GRADE: Graph Reconstruction via Additive Differential Equations

Our Proposal

GRADE: Graph Reconstruction via Additive Differential Equations

- Suppose ODE is additive

$$X_j'(t) = \sum_{k=1}^{p} f_{jk}\big(X_k(t)\big), \quad j = 1, \ldots, p$$

Our Proposal

GRADE: Graph Reconstruction via Additive Differential Equations

- Suppose ODE is additive

$$X_j'(t) = \sum_{k=1}^{p} f_{jk}\big(X_k(t)\big), \quad j = 1, \ldots, p$$

- Estimate $\hat{X}_j(t)$ from Y_j (using, e.g., nonparametric regression, etc)
- Calculate $\hat{\Psi}_{ik} = \int_0^{t_i} \psi(\hat{X}_k(s))\, ds, i = 1, \ldots, n$

Our Proposal

GRADE: Graph Reconstruction via Additive Differential Equations

- Suppose ODE is additive

$$X_j'(t) = \sum_{k=1}^{p} f_{jk}\big(X_k(t)\big), \quad j = 1, \ldots, p$$

- Estimate $\hat{X}_j(t)$ from Y_j (using, e.g., nonparametric regression, etc)
- Calculate $\hat{\Psi}_{ik} = \int_0^{t_i} \psi(\hat{X}_k(s))\, ds, i = 1, \ldots, n$
- Find the minimizer $\hat{\theta}_j$ of

$$\sum_{i=1}^{n} \Big[Y_{ij} - \underbrace{t_i \theta_{j0} - \sum_{i=1}^{n} \hat{\Psi}_{ik}^{\mathrm{T}} \theta_{jk}}_{\hat{f}_j} \Big]^2 + \lambda \sum_{k=1}^{p} \underbrace{\Big[\sum_{i=1}^{n} \big(\hat{\Psi}_{ik}^{\mathrm{T}} \theta_{jk} \big)^2 \Big]^{1/2}}_{\text{group lasso}}$$

Theory – I

- Variable selection consistency for

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

Theory – I

- Variable selection consistency for

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

- Replacing $\hat{\Psi}_{ik}$ with Ψ_{ik}, we have a standardized group lasso regression

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\Psi_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\Psi_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

  whose theoretical properties are well-understood

Theory – I

- Variable selection consistency for

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

- Replacing $\hat{\Psi}_{ik}$ with Ψ_{ik}, we have a standardized group lasso regression

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\Psi_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\Psi_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

  whose theoretical properties are well-understood
- With $\hat{\Psi}_{ik}$ we have an errors-in-variables regression

Theory – I

- Variable selection consistency for

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\hat{\Psi}_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

- Replacing $\hat{\Psi}_{ik}$ with Ψ_{ik}, we have a standardized group lasso regression

$$\sum_{i=1}^{n}\left[Y_{ij}-t_i\theta_{j0}-\sum_{i=1}^{n}\Psi_{ik}^{\mathrm{T}}\theta_{jk}\right]^2+\lambda\sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(\Psi_{ik}^{\mathrm{T}}\theta_{jk}\right)^2\right]^{1/2}$$

  whose theoretical properties are well-understood
- With $\hat{\Psi}_{ik}$ we have an errors-in-variables regression
- Need a bound on $\|\hat{\Psi}-\Psi\|$

Theory – II

Theory – II

- We establish a new concentration inequality to bound

$$\int_0^1 \left\{\hat{X}_j(t) - X_j(t)\right\}^2 dt$$

  - This inequality allows us to bound $\|\hat{\Psi} - \Psi\|$ in high dimensions, when $\log p/n^{\alpha} = o(1)$ for some $0 < \alpha < 0.5$
  - Using this inequality, the bound for derivative is asymptotically worst

Theory – II

- We establish a new concentration inequality to bound

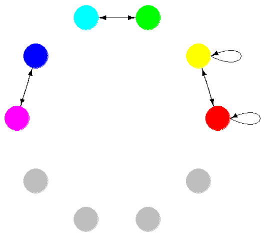
$$\int_0^1 \{\hat{X}_j(t) - X_j(t)\}^2 dt$$

  - This inequality allows us to bound $\|\hat{\Psi} - \Psi\|$ in high dimensions, when $\log p/n^{\alpha} = o(1)$ for some $0 < \alpha < 0.5$
  - Using this inequality, the bound for derivative is asymptotically worst

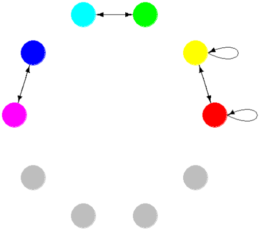
- We show that GRADE can consistently select the parents of each node in a sparse high-dimensional ODE network
  - The proof requires establishing model selection consistency of (standardized) group lasso regression with errors-in-variables[6]

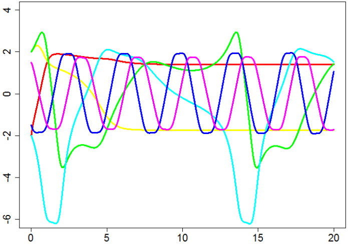
[6] Extending lasso (Loh & Wainwright, 2012 and Rosenbaum & Tsybakov, 2010)

Simulation: Design

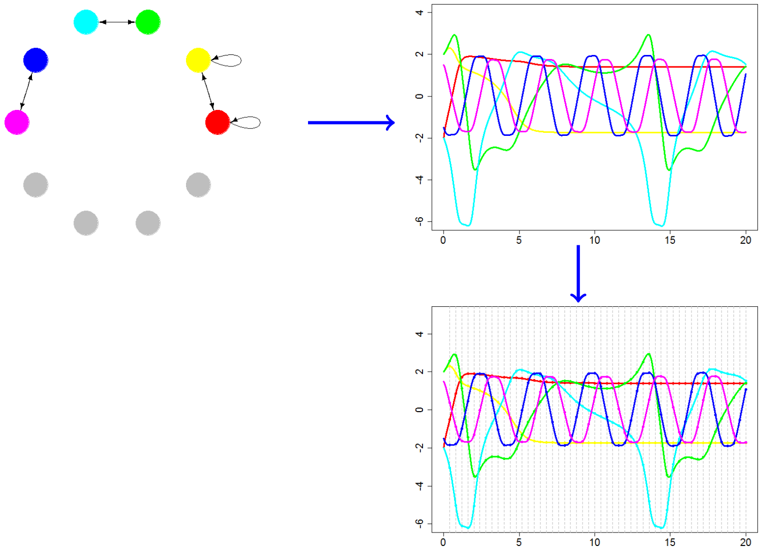


Simulation: Design

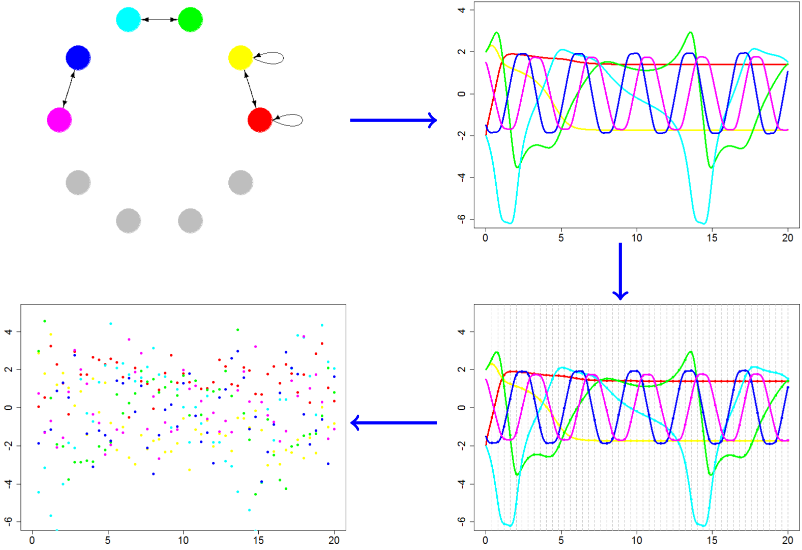




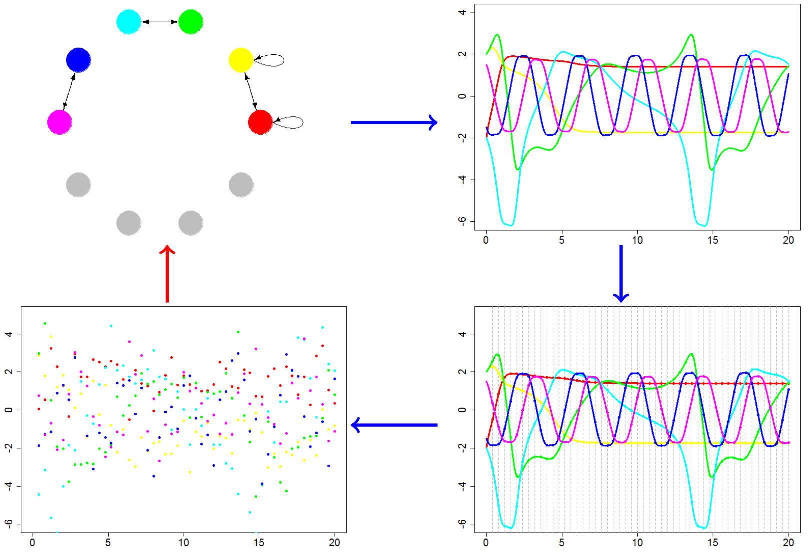
Simulation: Design



Simulation: Design

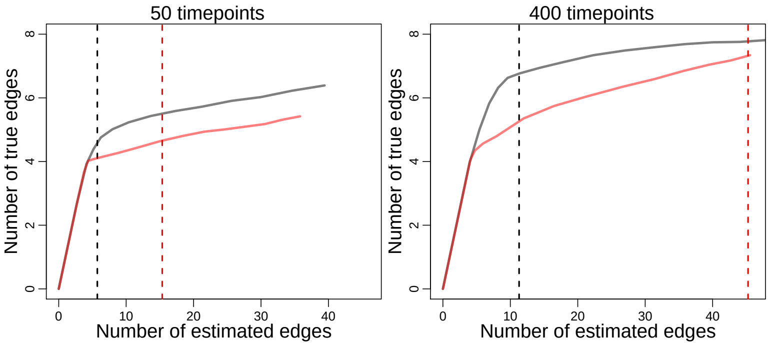


Simulation: Design



Simulation: Results

- NeRDS: Network Reconstruction via Dynamic Systems[7]
- GRADE



[7] Henderson & Michailidis (2014)

Application: DREAM-3 Challenge[8] (Regulatory Networks)

- *in silico* data from 5 regulatory networks with $p = 10$ or 100; $n = 50$
- A difficult task: non-additive ODEs with unobserved latent variables

[8] Schaffter et al (2011)

Application: DREAM-3 Challenge[8] (Regulatory Networks)

- *in silico* data from 5 regulatory networks with $p = 10$ or 100; $n = 50$
- A difficult task: non-additive ODEs with unobserved latent variables

Table: Area Under ROC Curves for NeRDS and GRADE

	$p = 10$		$p = 100$	
Network	NeRDS	GRADE	NeRDS	GRADE
Ecoli1	0.450	**0.545**	0.624	**0.670**
Ecoli2	0.512	**0.643**	0.637	**0.653**
Yeast1	0.486	**0.679**	0.610	**0.636**
Yeast2	0.525	**0.607**	0.568	**0.584**
Yeast3	0.467	**0.576**	**0.617**	0.567

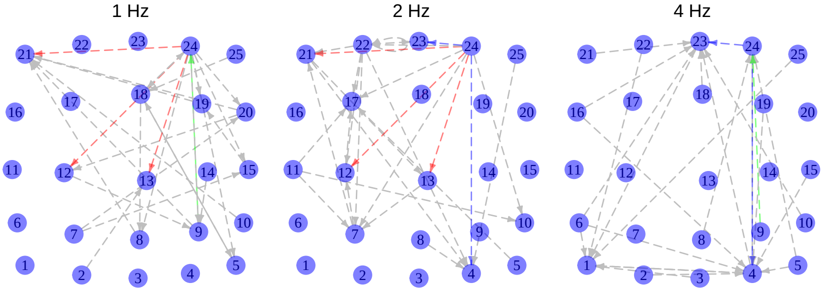
[8] Schaffter et al (2011)

Application: Brain Functional Connectivity

- *Cortical activity map (CAM) project* – Allen Institute for Brain Science
- Calcium fluorescent imaging in a region of visual cortex at 175mm depth measured using two-photons technology
- 575 neurons $\rightarrow$ 25 neuronal populations (5 $\times$ 5 grids with $\sim$20 neurons)
- 3 stimuli: frequencies of 1, 2, and 4 Hz, at a 90° spatial orientation
- $R = 15$ repetitions, $n = 60$ time points per repetition

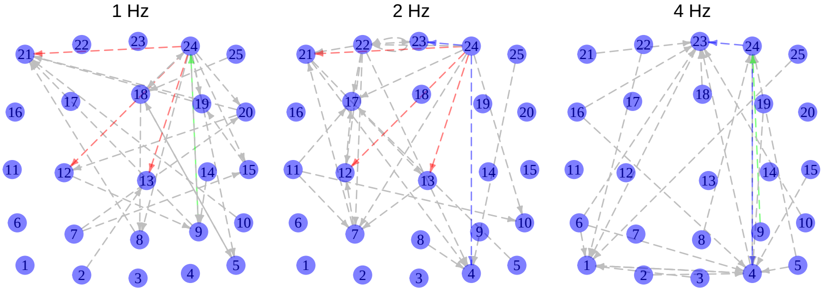
Application: Brain Functional Connectivity

- *Cortical activity map (CAM) project* – Allen Institute for Brain Science
- Calcium fluorescent imaging in a region of visual cortex at 175mm depth measured using two-photons technology
- 575 neurons $\rightarrow$ 25 neuronal populations (5 $\times$ 5 grids with $\sim$20 neurons)
- 3 stimuli: frequencies of 1, 2, and 4 Hz, at a 90° spatial orientation
- $R = 15$ repetitions, $n = 60$ time points per repetition



Application: Brain Functional Connectivity

- *Cortical activity map (CAM) project* – Allen Institute for Brain Science
- Calcium fluorescent imaging in a region of visual cortex at 175mm depth measured using two-photons technology
- 575 neurons $\rightarrow$ 25 neuronal populations (5 $\times$ 5 grids with $\sim$20 neurons)
- 3 stimuli: frequencies of 1, 2, and 4 Hz, at a 90° spatial orientation
- $R = 15$ repetitions, $n = 60$ time points per repetition



- More similar connectivity networks for closer frequencies

Summary

Summary

- Integration is better than differentiation!

Summary

- Integration is better than differentiation!
- GRADE takes advantage of the special structure of additive ODEs
  - It uses the linearity in parameters of truncated bases to avoid the estimation of derivatives
- Empirical & theoretical evidence shows improved performance
- GCV for bandwidth selection results in consistent estimates

Summary

- Integration is better than differentiation!

- GRADE takes advantage of the special structure of additive ODEs
  - It uses the linearity in parameters of truncated bases to avoid the estimation of derivatives
- Empirical & theoretical evidence shows improved performance
- GCV for bandwidth selection results in consistent estimates

- How can this idea be generalize to non-additive ODEs?

Acknowledgments:

- Grants from NSF-DMS & NIH-NIGMS
- Allen Institute for Brain Sciences for calcium imaging data and authors of existing methods for providing code

Reference:

- Chen, **S.** & Witten (2016), JASA, in press

Acknowledgments:

- Grants from NSF-DMS & NIH-NIGMS
- Allen Institute for Brain Sciences for calcium imaging data and authors of existing methods for providing code

Reference:

- Chen, **S.** & Witten (2016), JASA, in press

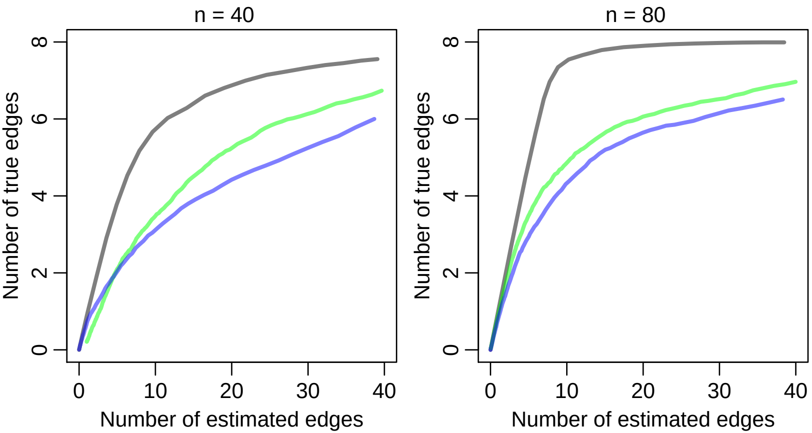
Thank You!

Comparison with Methods for Linear ODEs

- Linear ODEs of the form $X'(t) = \Theta X(t) + C$
- Comparison of GRADE with Hall & Ma (2014) and Brunel et al (2014)

Comparison with Methods for Linear ODEs

- Linear ODEs of the form $X'(t) = \Theta X(t) + C$
- Comparison of GRADE with Hall & Ma (2014) and Brunel et al (2014)



A Concentration Inequality

Theorem:
Under standard assumptions and if $\varepsilon_j, j = 1, \ldots, p$ are i.i.d. $N(0,1)$, the local polynomial regression estimator $\hat{X}(\cdot)$ satisfies

$$\int_0^1 \left\{\hat{X}_j(t) - X_j(t)\right\}^2 dt \le c_1 n^{\frac{2\beta}{2\beta+1}(\alpha - 0.5)}$$

for all $j = 1, \ldots, p$, with probability converging to 1 if

$$p \exp(-c_2 n^{2\alpha}) = o(1).$$

Remarks:

- Here, β and α are constants related to the smoothness of X and the choice of bandwidth for $\hat{X}$
- GCV, CV, and other methods can be used to choose the bandwidth
- For $\|\hat{X}' - X'\|_2$, the rate is $n^{\frac{2\beta-2}{2\beta-1}(\alpha-0.5)}$

Model Selection Consistency

Theorem:
Let

$$N_j^* = \{k : \|\theta_{jk}\|_2 \neq 0\}, \quad j = 1, \ldots, p$$

be the true parents of $X_j(\cdot)$ in the ODE network, and $\hat{N}_j$ be its estimator using the proposed method. Then, under certain regularity conditions, as the number of time points n increases,

$$\mathbb{P}\left(\hat{N}_j = N_j^* \text{ for all } j = 1, \ldots, p\right) \to 1.$$

Remark:

- The proof requires establishing model selection consistency of (standardized) group lasso regression with errors-in-variables