Network Reconstruction via Dynamic Systems:
Biological Network Reconstruction using Differential Equation Models

James Henderson and George Michailidis
jbhender@umich.edu, gmichail@umich.edu

University of Michigan
Statistics

Problem: Network Reconstruction

Network representations of biological systems are frequently used to provide a high-level description of the relationships among the system components. For instance, **regulatory networks** have nodes corresponding to biochemical complexes such as: genes, metabolites, proteins etc. A directed edge from node i to node j, $i \to j$, indicates a functional role for complex i in regulating the level of complex j. **Reconstructing the network** of regulatory relationships is a fundamental problem in systems biology.

Common Approaches

Conditional Independence Models, including *Bayesian Networks*, are based on learning a *directed acyclic graph* encoding conditional independence relations. *However, this approach cannot accomodate biologically relevant cycles such as feedback loops.* **Time Series Models** including *Dynamic Bayesian Networks* and *Vector Autoregressive Models* avoid this limitation by restricting edges to point forward in time. **ODE Models** are perhaps the most widespread class of models in the biological and physical sciences. *ODE-based methods* for network reconstruction from time-series data have been proposed but take an *ad-hoc* approach to inference and network reconstruction.

Networks, Dynamic Systems, & Coupling

This poster describes **methodology for network reconstruction from time series data, assumed to be noisy observations of an underlying dynamic system**,

$$\dot{x}^r(t) = f(\boldsymbol{x}^r(t)); \quad \boldsymbol{x}^r(0) = \boldsymbol{x}_0^r; \quad \boldsymbol{y}^r(t_k) = \boldsymbol{x}^r(t_k) + \boldsymbol{\epsilon}_k, \tag{1}$$

with r indexing experiments and the ϵ_k^r independent, mean zero, noise. **Only the trajectories are observed, not the derivatives. Network reconstruction is recast as variable selection for a dynamic system**; i.e. determining the important variables in f,

$$i \to j \iff \frac{\partial f_i}{\partial x_j}(\boldsymbol{x}(t)) \not\equiv 0. \tag{2}$$

Rank potential edges using the **coupling metric**,

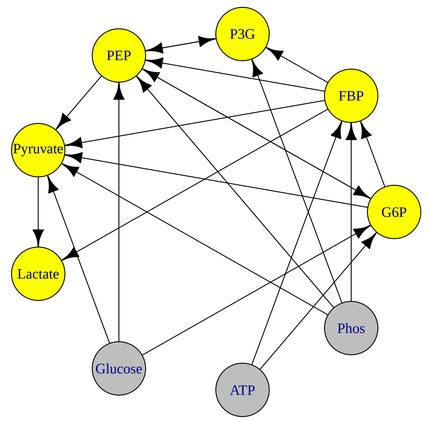
$$\rho(i,j) := \int \left| \frac{\partial f_i}{\partial x_j}(\boldsymbol{x}) \right| d\boldsymbol{x}, \tag{3}$$

measuring the strength of the regulatory relationship $i \to j$. Signed coupling metrics $\rho_+(i,j), \rho_-(i,j)$ can be defined similarly by replacing $|\cdot|$ with $(\cdot)_+$ or $(\cdot)_-$ respectively; useful for recovering signed edges.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. F032335 and No. F031344.

Case Study: *Lactococcus* Glucose Utilization



Embden-Meyerhof glycolitic pathway in Lactococcus Lactis. *Lactococcus Lactis is a lactic acid bacteria used in the industrial production of fermented milk products to convert glucose (or lactose) to lactic acid. The model described here is taken from (Voit, 2006). Yellow and grey nodes correspond to online and offline variables, respectively. Our goal is to reconstruct the network amont the online variables using time-series data but no other knowledge of the system.*

Setup

- Simulated data from a nonlinear ODE described in (Voit, 2006).
- Baseline initial values: $\boldsymbol{x}_0 = (2,2,2,2,1,1)'$
- We simulate perturbations of altering baseline initial values, $x_i^r(0) = M x_{0i}$
- Observe: $y_i^r(t_k) = x_i^r(t_k) + \epsilon_{ik}^r, \quad \epsilon_{ik}^r \sim N(0, (\sigma x_i^r(t_k))^2)$

Evaluation

		$\sigma = .02$		$\sigma = .05$	
		PR	ROC	PR	ROC
M = 10	Additive ODE	**0.90**	**0.89**	0.89	**0.87**
	Linear ODE	0.85	0.82	0.85	0.83
	Linear ODE + Lasso	0.67	0.54	0.66	0.55
M = 5	Additive ODE	**0.86**	**0.83**	**0.86**	**0.83**
	Linear ODE	0.77	0.74	0.77	0.75
	Linear ODE + Lasso	0.64	0.56	0.68	0.58
M = 2	Additive ODE	0.58	0.63	0.57	0.59
	Linear ODE	0.56	0.64	0.65	0.65
	Linear ODE + Lasso	0.50	0.59	0.56	0.61
M = 1.5	Additive ODE	0.49	0.57	0.46	0.51
	Linear ODE	**0.54**	**0.64**	0.57	0.65
	Linear ODE + Lasso	0.42	0.54	0.49	0.57

ODE Estimation: A Two-Stage Approach

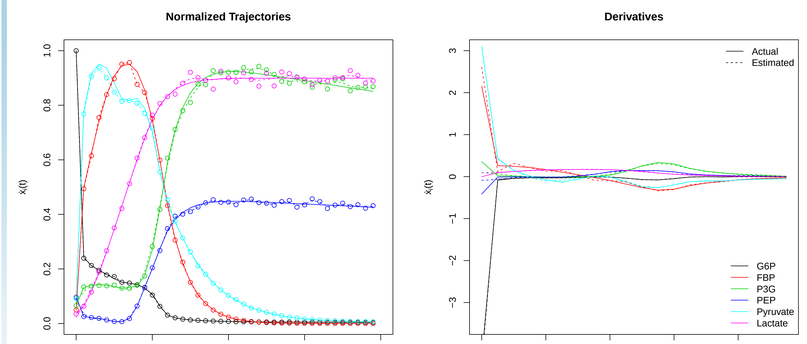
Stage 1: Estimate trajectories and derivatives using smoothing splines

- Avoids computationally intensive numerical integration
- Decouples the ODE estimation into d sub-problems

Stage 2: Fit an additive, non-parametric, approximation to the ODE

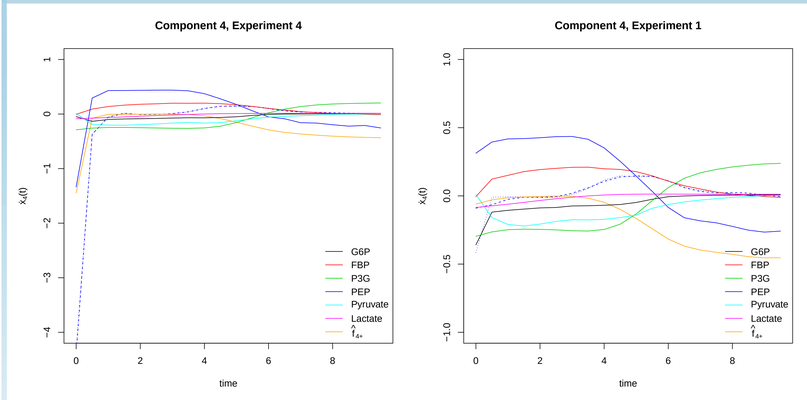
- The non-parametric approach allows for flexible nonlinear models and sidesteps potential model misspecification.
- Additive approximation for computational feasibility.

Stage 1: Normalize and Smooth



Trajectories, derivatives and their estimates in experiment $r = 1$.

Stage 2: Fit an Additive Model



The RHS function f is estimated by minimizing $\hat{f}_i = \arg\min \sum_{r=1}^R M_{in}^r(f)$,

$$M_{in}^r(f) = \int [\dot{\hat{x}}_i^r(t) - \sum_{j=1}^p f_{ij}(\hat{x}_i^r(t))]^2 dt + J(\lambda_1, \lambda_{2i}), \tag{4}$$

$$J(\lambda_1, \lambda_2) := \sum_{j=1}^p \lambda_1 \int [\ddot{f}_{ij}(x)]^2 dx + \lambda_2 \sum_{j=1}^p \sqrt{\int [f_{ij}(x)]^2 dx}. \tag{5}$$