Inferring Dynamics of Biological Systems

Tracey Oellerich[1]

In collaboration with: M. Emelianenko[1], M. Pierobon[2], E. Baldelli[2]

[1]Department of Mathematics, George Mason University, [2]Center for Applied Proteomics and Molecular Medicine, George Mason University

Abstract

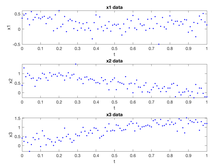
Biological networks are complex and constructing the dynamics for large systems can prove difficult. We will explore utilizing data-driven techniques to uncover the dynamics of a biological network from data provided. In particular, we are interested in inferring the dynamics for biological networks containing conservation laws or other special structures which can result in a singularity. These properties will be discussed in detail and their effect of recovering the dynamics will be investigated. Several numerical dynamics identification algorithms will be presented and their strengths and weaknesses will be discussed. In addition, the question of whether reconstructed networks exhibit adaptation properties based on recently developed analytical criteria will be investigated.

Methodology

We want to:

- Recover the dynamics from biological data

Given: Data from a biological network



Goal: Utilizing the data, recover a series of differential equations which accurately portray the data

$$x_1' = f_1(\mathbf{x})$$
$$x_2' = f_2(\mathbf{x})$$
$$x_3' = f_3(\mathbf{x})$$

where $\mathbf{x} = [x_1, x_2, x_3]$ is the vector of proteins of interest.

- Recreate network structure from dynamics
- Using recovered dynamics, test for properties of interest such as:
  - Conservation Laws [1]
  - Special network structures, such as compounds in the network
  - Response to input/stimuli
  - Adaptation properties [1]

Known problems:

- Rational functions are typically present in biological models (Michaelis–Menten kinetics, etc.)
- Systems can contain properties which could interfere with recovering dynamics
- May not have access to derivative information or it may not be accurate enough

Funding

This research is part of the Industrial Immersion Program funded by the George Mason Provost. Both AWM and AMS have provided travel funding to present this work.

Inferring Dynamics

We begin by examining the structure of the Sparse Identification of Nonlinear Dynamics (SINDy) Method [2]:
Consider a dynamical system:

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x})$$

where we wish to approximate $\mathbf{f}(\mathbf{x})$ by a generalized linear model:

$$\mathbf{f}_k(\mathbf{x}) \approx \Theta(\mathbf{x})\xi_k$$

with the fewest nonzero terms in ξ_k as possible.
Let $\mathbf{X} = \begin{bmatrix}\mathbf{x}(t_1) & \dots & \mathbf{x}(t_m)\end{bmatrix}^T$ be the time series data collected from the system. Then, we can develop a library of candidate nonlinear functions $\Theta(\mathbf{X})$ constructed from the data in $\mathbf{X}$:

$$\Theta(\mathbf{X}) = \begin{bmatrix}\mathbf{1} & \mathbf{X} & \mathbf{X}^2 & \dots & \mathbf{X}^d & \dots & \sin(\mathbf{X}) & \dots\end{bmatrix}$$

Now the dynamical system can be represented in the form:

$$\dot{\mathbf{X}} = \begin{bmatrix}\dot{\mathbf{x}}(t_1) & \dots & \dot{\mathbf{x}}(t_m)\end{bmatrix}^T = \Theta(\mathbf{X})\begin{bmatrix}\xi_1 & \dots & \xi_k & \dots\end{bmatrix}$$

We wish to minimize:

$$\xi_k = \underset{\xi_k'}{\arg\min}\|\dot{\mathbf{X}}_k - \Theta(\mathbf{X})\xi_k'\|_2 + \lambda\|\xi_k'\|_1$$

where λ weights the sparsity constraint.

Alternative Framework: For our purposes, we will use the ideas presented in the SINDy Method and apply them utilizing a different solver.
Consider the dynamical system

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x})$$

for which we would like to recover from data provided.
We will consider an approximation of $\mathbf{f}(\mathbf{x})$ using a generalized linear model (as in the SINDy Method):

$$\mathbf{f}_k(\mathbf{x}) \approx \begin{bmatrix}\Theta(\mathbf{x}) \\ -\Theta(\mathbf{x})\end{bmatrix}\omega_k$$

where Θ is a library of candidate nonlinear functions constructed from the data and ω_k contains the fewest nonzero terms possible.
In order to find the optimal ξ_k, we will utilize the time series data in the **Non-negative Least Squares (NNLS)** algorithm [3] as follows:

$$\omega_k = \underset{\omega_k', \omega_k' \geq 0}{\arg\min}\left\|\begin{bmatrix}\Theta(\mathbf{X}) \\ -\Theta(\mathbf{X})\end{bmatrix}\omega_k' - \dot{\mathbf{X}}_k\right\|_2$$

where the top entries of ω_k will correspond to positive coefficients in the recovered dynamics and the bottom entries are the negative.

Inferring Dynamics with Rational Functions

We wish to discover the dynamics of:

$$\frac{d}{dt}x_k(t) = \frac{f_{N,k}(\mathbf{x})}{f_{D,k}(\mathbf{x})}$$

from time series data of the state $\mathbf{x}(t) = [x_1(t), ..., x_n(t)]^T$ and where $f_{N,k}(\mathbf{x})$ and $f_{D,k}(\mathbf{x})$ represent the numerator and denominator polynomials in the state variable $\mathbf{x}_k$.
We will proceed as in Mangan et al. [4] by multiplying both sides by the denominator:

$$f_{N,k}(\mathbf{x}) - f_{D,k}(\mathbf{x})\dot{x}_k = 0$$

As before, we wish to approximate $f_{N,k}(\mathbf{x})$ and $f_{D,k}(\mathbf{x})$ by generalized linear models:

$$f_{N,k} \approx \Theta_N(\mathbf{x})\omega_{N,k} \qquad f_{D,k} \approx \Theta_D(\mathbf{x})\omega_{D,k}$$

where $\Theta_N(\mathbf{x}), \Theta_D(\mathbf{x})$ are the candidate function libraries and $\omega_{N,k}, \omega_{D,k}$ are the corresponding coefficients.
We will consider the library of candidate linear functions constructed from the data:

$$\Theta_N(\mathbf{X}) = \Theta_D(\mathbf{X}) = \begin{bmatrix}\mathbf{1} & \mathbf{X} & \mathbf{X}^2 & \dots & \mathbf{X}^d & \dots\end{bmatrix}$$

Thus, we have:

$$\Theta_N(\mathbf{x})\omega_{N,k} - \Theta_D(\mathbf{x})\omega_{D,k}\dot{x}_k = 0$$

In order to apply NNLS, we will assume the coefficient associated with the $x_k\dot{x}_k$ term is 1 and thus solve:

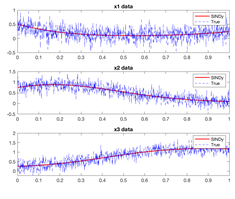
$$x_k\dot{x}_k = \Theta_N(\mathbf{x})\omega_{N,k} - \tilde{\Theta}_D(\mathbf{x})\omega_{D,k}\dot{x}_k$$

where $\tilde{\Theta}_D$ is the Θ_D matrix with the column corresponding to x_k removed. The optimal $\omega_{N,k}$ and $\omega_{D,k}$ can now be found using a similar method as before.

Recover Dynamics

$$\dot{x}_1 = k_3x_1x_3 - k_1x_1x_2$$
$$\dot{x}_2 = k_1x_1x_2 - k_2x_2x_3$$
$$\dot{x}_3 = k_2x_2x_3 - k_3x_1x_3$$
$$C_{tot} = x_1 + x_2 + x_3$$

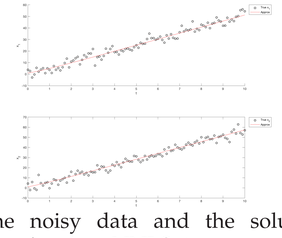
$$k_1 = 6, k_2 = 4, k_3 = 3$$
$$\mathbf{x}_0 = (.5, .75, .25)$$



The noisy data and the solution found using SINDy. Similar results were obtained using NNLS.

Recover Rational Functions

$$\frac{dx_1}{dt} = \frac{2x_1 + 3x_2}{1 + x_1}$$
$$\frac{dx_2}{dt} = \frac{2 + x_1 + 5x_2}{1 + x_2}$$



The noisy data and the solution found using NNLS to recover dynamics for systems with rational functions.

The MAPK Pathway

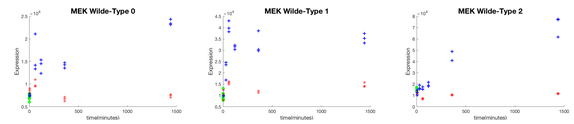
Above is a diagram for the MAPK Pathway provided by Dr. Pierobon.

Properties:

- Plays a role in regulation of gene expression and cellular growth
- Dysregulated MAPK signaling is implicated in a wide range of cancers
- Ras protein is prone to mutation
- Data for the proteins in this pathway is a measure of the protein activation

Data Properties

- Data from 8 different cell lines over a period of 24 hours with two different solutions applied, Dimethyl sulfoxide (DMSO) and the MEK Inhibitor Selumetinib.
- 3 different states for the KRAS protein:
  - Wilde-Type = 0: 0 normal copies of gene (mutated)
  - Wilde-Type = 1: 1 normal copies of gene (mutated)
  - Wilde-Type = 2: 2 normal copies of gene (**un**mutated)



MEK protein expression over time where o represent baseline value, * are DMSO Control, and + are Selumetinib (MEK Inhibitor).

Future Research

There are several avenues for future research:

- Recover a model for the MAPK pathway
  - Model differences between DMSO application and MEK inhibitor application
  - Model differences for mutated cell lines
  - For all models, identify any interesting properties of the dynamics, such as conservation laws, properties of adaptation, etc.
- Continue to develop the theory behind the recovery process
- What are the limitations for recovery?

References

[1] T. Oellerich, M. Emelianenko, L. A. Liotta, and R. P. Araujo. Biological networks with singular jacobians: their origins and adaptation criteria. *Submitted*.

[2] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[3] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. SIAM, 1995.

[4] N.M. Mangan, S.L. Brunton, J.L. Proctor, and J.N. Kutz. Inferring biological networks by sparse identification of nonlinear dynamics. *EEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2, 2016.