Learning sparse dynamics from limited measurements

Rachel Ward

February, 2019

Learning Governing Equations: Problem Set-up

- Unknown nonlinear dynamical system

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad \text{where } \mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d \text{ continuous.}$$

- Given: (possibly noisy or corrupted) limited snapshots of the system

$$\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m), \qquad t_m = m\Delta t$$

- Goal: Quickly recover governing equations $\mathbf{f} = (f_1, f_2, \ldots, f_d)$ from snapshots

- Problem is ill-posed without additional assumptions.

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad \text{where } \mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d \text{ continuous.}$$

- Reasonable assumption: Suppose $\mathbf{f} = (f_1, f_2, \ldots, f_d)$ are sparse, multivariate polynomials of maximal degree L to capture many dynamical systems of interest:

$$f_k(\mathbf{x}) = \sum_{\alpha_1 + \cdots + \alpha_d \leq L} c_\alpha^k x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d}$$

- Goal is then to determine polynomial coefficients c_α^k from data $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)$.
- Sparsity assumption: At most $s \ll d^L$ among each set of coefficients $\{c_\alpha^k\}, \quad k = 1, 2, \ldots, d$ are non-zero, we just don't know which ones are nonzero.

"Lifting Trick" to linearize problem

Form data and (approximate) velocity matrices from given snapshots: [1]

$$X = \begin{bmatrix} | & \cdots & | \\ X_1 & \cdots & X_d \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} x_1(t_1) & \cdots & x_d(t_1) \\ \vdots & \cdots & \vdots \\ x_1(t_m) & \cdots & x_d(t_m) \end{bmatrix}_{m \times d}, \quad \dot{X} = \begin{bmatrix} | & \cdots & | \\ \dot{X}_1 & \cdots & \dot{X}_d \\ | & \cdots & | \end{bmatrix}_{m \times d}$$

[1] Popularized by Brunton, Proctor, and Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *PNAS* 2016.

"Lifting Trick" to linearize problem

Form data and (approximate) velocity matrices from given snapshots: [1]

$$X = \begin{bmatrix} | & \cdots & | \\ X_1 & \cdots & X_d \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} x_1(t_1) & \cdots & x_d(t_1) \\ \vdots & \cdots & \vdots \\ x_1(t_m) & \cdots & x_d(t_m) \end{bmatrix}_{m \times d}, \quad \dot{X} = \begin{bmatrix} | & \cdots & | \\ \dot{X}_1 & \cdots & \dot{X}_d \\ | & \cdots & | \end{bmatrix}_{m \times d}$$

Construct *dictionary matrix* from data:

$$D_X = \begin{bmatrix} | & | & & | & | & | & & | & \\ 1 & X_1 & \cdots & X_d & X_1^2 & X_1 X_2 & \cdots & X_d^2 & \cdots \\ | & | & & | & | & | & & | & \end{bmatrix}_{m \times N}$$

[1] Popularized by Brunton, Proctor, and Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *PNAS* 2016.

"Lifting Trick" to linearize problem

Form data and (approximate) velocity matrices from given snapshots: [1]

$$X = \begin{bmatrix} | & \cdots & | \\ X_1 & \cdots & X_d \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} x_1(t_1) & \cdots & x_d(t_1) \\ \vdots & \cdots & \vdots \\ x_1(t_m) & \cdots & x_d(t_m) \end{bmatrix}_{m\times d}, \quad \dot{X} = \begin{bmatrix} | & \cdots & | \\ \dot{X}_1 & \cdots & \dot{X}_d \\ | & \cdots & | \end{bmatrix}_{m\times d}$$

Construct *dictionary matrix* from data:

$$D_X = \begin{bmatrix} | & | & & | & | & | & & | & \\ 1 & X_1 & \cdots & X_d & X_1^2 & X_1X_2 & \cdots & X_d^2 & \cdots \\ | & | & & | & | & | & & | & \end{bmatrix}_{m\times N}$$

Recover $\mathcal{C} = [c_\alpha^1, c_\alpha^2, \ldots c_\alpha^d]_{|\alpha|\le L} \in \mathbb{R}^{N\times d}$ as solution to the linear inverse problem

$$\dot{X} = D_X\,\mathcal{C} \qquad \text{or} \qquad \dot{X} = D_X\,\mathcal{C} + \mathcal{E}$$

where $\mathcal{E}$ is error in time-derivative approximations and $\|\cdot\|$ is the max of column ℓ_2 norms.

[1] Popularized by Brunton, Proctor, and Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *PNAS* 2016.

Algorithm for sparse reconstruction

We are interested in recovering $\mathcal{C}$ from $\dot{X}$ and D_X when number of measurements m is limited — $D_X \in \mathbb{R}^{m \times N}$ is underdetermined. Natural optimization algorithm to pick sparsest solution consistent with measurements:

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_0, \quad \text{s.t} \quad \|\dot{X} - D_X \mathcal{Z}\| \leq \sigma$$

Algorithm for sparse reconstruction

We are interested in recovering $\mathcal{C}$ from $\dot{X}$ and D_X when number of measurements m is limited — $D_X \in \mathbb{R}^{m \times N}$ is underdetermined. Natural optimization algorithm to pick sparsest solution consistent with measurements:

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_0, \quad \text{s.t} \quad \|\dot{X} - D_X \mathcal{Z}\| \leq \sigma$$

The sparse optimization algorithm above is NP hard, so relax to an L_1 algorithm (convex, called Basis Pursuit):

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_1, \quad \text{s.t} \quad \|\dot{X} - D_X \mathcal{Z}\| \leq \sigma$$

Algorithm for sparse reconstruction

We are interested in recovering $\mathcal{C}$ from $\dot{X}$ and D_X when number of measurements m is limited — $D_X \in \mathbb{R}^{m \times N}$ is underdetermined. Natural optimization algorithm to pick sparsest solution consistent with measurements:

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_0, \quad \text{s.t} \quad \|\dot{X} - D_X \mathcal{Z}\| \leq \sigma$$

The sparse optimization algorithm above is NP hard, so relax to an L_1 algorithm (convex, called Basis Pursuit):

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_1, \quad \text{s.t} \quad \|\dot{X} - D_X \mathcal{Z}\| \leq \sigma$$

Approach: Random initialization, apply results from compressive sensing: Properties of D_X such that in the range $m \ll N$, the solutions of the two problems agree, or are close.

Sparse Recovery from Multiple Trajectories

- Suppose we can collect snapshots from K different trajectories:

$$\{\mathbf{x}(t_1, 1), \ldots, \mathbf{x}(t_m, 1)\}, \quad \{\mathbf{x}(t_1, K), \ldots, \mathbf{x}(t_m, K)\},$$

- Form $\dot{X}$ and dictionary matrix D_X of size $mK \times N$ and solve for $\mathcal{C}$.

Sparse Recovery from Multiple Trajectories

- Suppose we can collect snapshots from K different trajectories:

$$\{\mathbf{x}(t_1, 1), \ldots, \mathbf{x}(t_m, 1)\}, \quad \{\mathbf{x}(t_1, K), \ldots, \mathbf{x}(t_m, K)\},$$

- Form $\dot{X}$ and dictionary matrix D_X of size $mK \times N$ and solve for $\mathcal{C}$.

Theorem (Schaeffer, Tran, and W. 2017)

Assume each component of $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_d(\mathbf{x}))$ *is a multivariate polynomial of maximal order* L*, and has at most* s *non-zero polynomial coefficients – call this coefficient matrix* $\mathcal{C}$*. Assume the* K *initializations* $\{\mathbf{x}(t_1, 1), \ldots, \mathbf{x}(t_1, K)\}$ *are drawn i.i.d. uniformly from* $[-1, 1]^d$*; and that* $K \geq c\, 3^L s \log(d) \log(\varepsilon^{-1})$.

Sparse Recovery from Multiple Trajectories

- Suppose we can collect snapshots from K different trajectories:

$$\{\mathbf{x}(t_1,1),\ldots,\mathbf{x}(t_m,1)\},\quad \{\mathbf{x}(t_1,K),\ldots,\mathbf{x}(t_m,K)\},$$

- Form $\dot{X}$ and dictionary matrix D_X of size $mK \times N$ and solve for $\mathcal{C}$.

Theorem (Schaeffer, Tran, and W. 2017)

Assume each component of $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}),\ldots,f_d(\mathbf{x}))$ *is a multivariate polynomial of maximal order* L*, and has at most* s *non-zero polynomial coefficients – call this coefficient matrix* $\mathcal{C}$*. Assume the* K *initializations* $\{\mathbf{x}(t_1,1),\ldots,\mathbf{x}(t_1,K)\}$ *are drawn i.i.d. uniformly from* $[-1,1]^d$*; and that* $K \geq c\,3^L s\,\log(d)\log(\varepsilon^{-1})$.

Then with probability $1-\varepsilon$*,* $\mathcal{C}$ *is the unique solution to the* ℓ_1*-minimization problem:*

$$\min \|\mathcal{Z}\|_1 \quad \textit{subject to} \quad \dot{X} = D_X\,\mathcal{Z},$$

and recovery is stable with respect to inexact sparsity and robust with respect to additive noise (as from approximating derivatives).

Sparse Recovery from Multiple Trajectories

- Reconstruction guarantees can be extended to sparsity with respect to other bounded orthonormal bases such as sines and cosines, but not to an arbitrary dictionary (incoherence required).
- Our recovery guarantee uses only the initial measurements from each burst; later measurements used only the approximate $\dot{\mathbf{x}}(t_1, \ell)$. Compressive sensing guarantees require randomness/ Central Limit Theorems for sampling points.
- [Tran, Ward 2016] If $\{\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)\}$ is sufficiently *ergodic*, then the iterates act similar to i.i.d. sampling points (central limit theorem), and one can derive guarantees using measurements from a single trajectory.
- Can we steer the system towards an ergodic trajectory to reduce number of measurements?