Discovering governing equations using noisy measurements through projection-based denoising and second order cone programming

Jacqueline Wentz, Alireza Doostan

University of Colorado at Boulder
Machine Learning/Deep Learning Conference, July 2022

Funding: DOE Predictive Science Academic Alliance Program (PSAAP)
Discover governing equations by assuming basis expansion

- Goal is to find governing equations of an ODE system

\[
\begin{align*}
\dot{u}_1 &= c_{1,1} + c_{1,2}u_1 + c_{1,3}u_2 + c_{1,4}u_1u_2 + \ldots \\
\dot{u}_2 &= c_{2,1} + c_{2,2}u_1 + c_{2,3}u_2 + c_{2,4}u_1u_2 + \ldots
\end{align*}
\]

Suppose we have \( N \) state measurements \( u_1(t_i) \) and \( u_2(t_i) \) for \( i = 1, 2, \ldots, N \)

Find \( c_i \) for \( i = 1, 2 \) such that

\[
\Theta c_i = \dot{u}_i, \quad i = 1, 2
\]

where

\[
\Theta = \begin{bmatrix}
1 & u_1(t_1) & u_2(t_1) & \cdots & 1 & u_1(t_2) & u_2(t_2) & \cdots & 1 & u_1(t_N) & u_2(t_N) & \cdots
\end{bmatrix}
\]

and

\[
\dot{u}_i = \begin{bmatrix}
\dot{u}_i(t_1) & \dot{u}_i(t_2) & \cdots & \dot{u}_i(t_N)
\end{bmatrix}
\]
Discover governing equations by assuming basis expansion

- Goal is to find governing equations of an ODE system
- In system with two states $u_1$ and $u_2$, for example, we first assume:

$$
\dot{u}_1 = c_{1,1} + c_{1,2}u_1 + c_{1,3}u_2 + c_{1,4}u_1^2 + c_{1,5}u_1u_2 + ... \\
\dot{u}_2 = c_{2,1} + c_{2,2}u_1 + c_{2,3}u_2 + c_{2,4}u_2^2 + c_{2,5}u_1u_2 + ...
$$

(1)
Discover governing equations by assuming basis expansion

- Goal is to find governing equations of an ODE system
- In system with two states \( u_1 \) and \( u_2 \), for example, we first assume:

\[
\begin{align*}
\dot{u}_1 &= c_{1,1} + c_{1,2}u_1 + c_{1,3}u_2 + c_{1,4}u_1^2 + c_{1,5}u_1u_2 + \ldots \\
\dot{u}_2 &= c_{2,1} + c_{2,2}u_1 + c_{2,3}u_2 + c_{2,4}u_1^2 + c_{2,5}u_1u_2 + \ldots
\end{align*}
\] (1)

- Suppose we have \( N \) state measurements \( \rightarrow u_1(t_i) \) and \( u_2(t_i) \) for \( i = 1, 2, \ldots, N \)
Discover governing equations by assuming basis expansion

- Goal is to find governing equations of an ODE system
- In system with two states \( u_1 \) and \( u_2 \), for example, we first assume:

\[
\begin{align*}
\dot{u}_1 &= c_{1,1} + c_{1,2}u_1 + c_{1,3}u_2 + c_{1,4}u_1^2 + c_{1,5}u_1u_2 + \ldots \\
\dot{u}_2 &= c_{2,1} + c_{2,2}u_1 + c_{2,3}u_2 + c_{2,4}u_1^2 + c_{2,5}u_1u_2 + \ldots
\end{align*}
\]  

(1)

- Suppose we have \( N \) state measurements \( \rightarrow u_1(t_i) \) and \( u_2(t_i) \) for \( i = 1, 2, \ldots, N \)
- Find \( c_i \) for \( i = 1, 2 \) such that

\[
\Theta c_i = \dot{u}_i, \quad \text{for } i = 1, 2
\]

(2)

where

\[
\Theta = \begin{bmatrix}
1 & u_1(t_1) & u_2(t_1) & u_1^2(t_1) & \ldots \\
1 & u_1(t_2) & u_2(t_2) & u_1^2(t_2) & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
1 & u_1(t_N) & u_2(t_N) & u_1^2(t_N) & \ldots
\end{bmatrix}, \quad \dot{u}_i = \begin{bmatrix}
\dot{u}_i(t_1) \\
\dot{u}_i(t_2) \\
\vdots \\
\dot{u}_i(t_N)
\end{bmatrix}
\]

(3)
Existing approaches for coefficient recovery fail at large noise levels

- **Sparse identification of nonlinear dynamics (SINDy):** Brunton et al. 2016
  1. Estimate the derivative, $\dot{u}$, using measurements
  2. Assume coefficient vector is sparse
  3. Solve the following, using sequential thresholding least squares:

\[
\mathbf{c} = \arg \min_{\mathbf{c}'} \| \Theta \mathbf{c}' - \dot{u} \|_2^2 + \lambda \| \mathbf{c}' \|_1
\]  

(4)

- Challenges:
  1. Derivative estimation in the presence of noise
  2. Determining optimal $\lambda$

- Research Question: Can we improve coefficient recovery in the presence of noise?
  1. Novel denoising strategy
  2. Approach for finding the coefficients/derivative simultaneously.
Existing approaches for coefficient recovery fail at large noise levels

  1. Estimate the derivative, $\dot{u}$, using measurements
  2. Assume coefficient vector is sparse
  3. Solve the following, using sequential thresholding least squares:

$$ c = \arg \min_{c'} \| \Theta c' - \dot{u} \|_2^2 + \lambda \| c' \|_1 $$

(4)

- Challenges:
  1. Derivative estimation in the presence of noise
  2. Determining optimal $\lambda$
Existing approaches for coefficient recovery fail at large noise levels


1. Estimate the derivative, $\dot{u}$, using measurements
2. Assume coefficient vector is sparse
3. Solve the following, using sequential thresholding least squares:

$$c = \arg \min_{c'} \| \Theta c' - \dot{u} \|_2^2 + \lambda \| c' \|_1$$

(4)

Challenges:

1. Derivative estimation in the presence of noise
2. Determining optimal $\lambda$

Research Question: Can we improve coefficient recovery in the presence of noise?

1. Novel denoising strategy
2. Approach for finding the coefficients/derivative simultaneously.
Example dynamic system: Van der Pol oscillator

▶ Non-conservative oscillator with non-linear damping:

\[ \frac{d^2 x}{dt^2} - \mu(1 - x^2)\frac{dx}{dt} + x = 0. \]  \hspace{1cm} (5)

▶ Equivalent ODE:

\[ x \rightarrow u_1 \text{ and } \dot{x} \rightarrow u_2 \]

\[ \dot{u}_1 = u_2, \quad \dot{u}_2 = -u_1 + \mu u_2 - \mu u_1 u_2 \]  \hspace{1cm} (6)

▶ Example dynamics:

\[ \mu = 2, \quad u(0) = [0, 1] \]

▶ Noise \( \epsilon \sim \mathcal{N}(0, \sigma^2) \), e.g., \( \sigma^2 = 0.1 \).
Example dynamic system: Van der Pol oscillator

- Non-conservative oscillator with non-linear damping:

\[
\frac{d^2 x}{dt^2} - \mu (1 - x^2) \frac{dx}{dt} + x = 0.
\] (5)

- Equivalent ODE: \( x \rightarrow u_1 \) and \( \dot{x} \rightarrow u_2 \)

\[
\begin{align*}
\dot{u}_1 &= u_2 \\
\dot{u}_2 &= -u_1 + \mu u_2 - \mu u_1^2 u_2
\end{align*}
\] (6)
Example dynamic system: Van der Pol oscillator

- Non-conservative oscillator with non-linear damping:

\[
\frac{d^2x}{dt^2} - \mu(1 - x^2)\frac{dx}{dt} + x = 0. \quad (5)
\]

- Equivalent ODE: \( x \rightarrow u_1 \) and \( \dot{x} \rightarrow u_2 \)

\[
\dot{u}_1 = u_2 \\
\dot{u}_2 = -u_1 + \mu u_2 - \mu u_1^2 u_2 \quad (6)
\]

- Example dynamics: \( \mu = 2, \ u^{(0)} = [0, 1] \)
Example dynamic system: Van der Pol oscillator

- Non-conservative oscillator with non-linear damping:
  \[
  \frac{d^2 x}{dt^2} - \mu (1 - x^2) \frac{dx}{dt} + x = 0. \tag{5}
  \]

- Equivalent ODE: \( x \rightarrow u_1 \) and \( \dot{x} \rightarrow u_2 \)
  \[
  \dot{u}_1 = u_2 \\
  \dot{u}_2 = -u_1 + \mu u_2 - \mu u_1^2 u_2 \tag{6}
  \]

- Example dynamics: \( \mu = 2, \ u^{(0)} = [0, 1] \)

- Noise \( \epsilon \sim \mathcal{N}(0, \sigma^2) \), e.g., \( \sigma^2 = 0.1 \).
Projection-based state denoising: Method

- Use the assumed basis given by the columns of $\Theta$. 

\[ \Theta_c = \dot{u}, \quad (7) \]

\[ \Phi_d = u \quad \text{where} \quad \Phi = [1^T \Theta] \]

\[ \tilde{u} = P_{\Phi} u \quad (9) \]

This implies we expect $u$ to be in the column space of $\Theta$. 

$P_{\Phi}$ is projection operator calculated using $u$. 
Projection-based state denoising: Method

- Use the assumed basis given by the columns of $\Theta$.
- We apply quadrature techniques to

$$\Theta c = \dot{u},$$  \hspace{1cm} (7)

to obtain,

$$\Phi d = u \quad \text{where} \quad \Phi = \begin{bmatrix} 1 & T\Theta \end{bmatrix}.$$  \hspace{1cm} (8)

This implies we expect $u$ to be in the column space of $\Phi$.

Project data onto this expected subspace:

$$\tilde{u} = P\Phi u$$  \hspace{1cm} (9)

where $P\Phi$ is projection operator calculated using $u$. 

Projection-based state denoising: Method

- Use the assumed basis given by the columns of $\Theta$.
- We apply quadrature techniques to

$$\Theta c = \dot{u}, \quad (7)$$

to obtain,

$$\Phi d = u \quad \text{where} \quad \Phi = \begin{bmatrix} 1 & T\Theta \end{bmatrix}. \quad (8)$$

- This implies we expect $u$ to be in the column space of $\Phi$. 
Projection-based state denoising: Method

- Use the assumed basis given by the columns of $\Theta$.
- We apply quadrature techniques to

$$\Theta c = \dot{u},$$

(7)

to obtain,

$$\Phi d = u \quad \text{where} \quad \Phi = \begin{bmatrix} 1 & T\Theta \end{bmatrix}.$$  

(8)

- This implies we expect $u$ to be in the column space of $\Phi$.
- Project data onto this expected subspace:

$$\tilde{u} = P_\Phi u$$

(9)

where $P_\Phi$ is projection operator calculated using $u$. 
Projection-based denoising: Results for Van der Pol oscillator

Relative $\ell^2$ error $u_1$

- **Projection Method**
- **Gaussian Process**

Noise level ($\sigma$)

$10^{-3}$ $10^{-2}$ $10^{-1}$ $10^{0}$

Relative $\ell^2$ error $u_2$

- **Projection Method**
- **Gaussian Process**

Noise level ($\sigma$)

$10^{-3}$ $10^{-2}$ $10^{-1}$ $10^{0}$
Learn coefficients and derivative simultaneously

- Write coefficients in terms of derivative:

\[
\Theta c = \dot{u} \quad \rightarrow \quad c = (\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T \dot{u}.
\] (10)
Learn coefficients and derivative simultaneously

- Write coefficients in terms of derivative:

\[ \Theta c = \dot{u} \quad \rightarrow \quad c = (\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T \dot{u}. \] (10)

- Set \( B \) as estimator of \( (\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T \).

- Second order cone program (SOCP) to find initial condition, \( u_0 \), and derivative, \( \dot{u} \):

\[
\begin{align*}
\text{minimize} & \quad u_0, \quad \dot{u} \\
\text{subject to} & \quad \|D \dot{u}\|_2 \leq C \quad \text{(smooth derivative)} \\
& \quad \|\left[1^T, u_0, \dot{u}\right] - \tilde{u}\|_2 \leq \gamma \quad \text{(match a priori smoothed data)}.
\end{align*}
\] (11)

- Estimate \( C \) using projection-based denoising result.

- Find \( \gamma \) using the corner point of Pareto curve.
Learn coefficients and derivative simultaneously

- Write coefficients in terms of derivative:
  \[
  \Theta c = \dot{u} \quad \Rightarrow \quad c = (\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T \dot{u}.
  \]  
  (10)

- Set \( B \) as estimator of \((\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T\).

- Second order cone program (SOCP) to find initial condition, \( u_0 \), and derivative, \( \dot{u} \):
  \[
  \begin{aligned}
  \text{minimize}_{u_0, \dot{u}} & \quad \|B\dot{u}\|_1 & \quad \text{(sparsity of coefficients)} \\
  \text{subject to} & \quad \|D\dot{u}\|_2 \leq C & \quad \text{(smooth derivative)} \\
  & \quad \left\| \begin{bmatrix} 1 & T \end{bmatrix} \begin{bmatrix} u_0 \\ \dot{u} \end{bmatrix} - \tilde{u} \right\|_2 \leq \gamma & \quad \text{(match a priori smoothed data)}. 
  \end{aligned}
  \] 
  (11)

- Estimate \( C \) using projection-based denoising result.

- Find \( \gamma \) using the corner point of Pareto curve.
Learn coefficients and derivative simultaneously

- Write coefficients in terms of derivative:

$$\Theta c = \dot{u} \rightarrow c = (\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T \dot{u}.$$  \hspace{1cm} (10)

- Set $B$ as estimator of $(\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T$.

- Second order cone program (SOCP) to find initial condition, $u_0$, and derivative, $\dot{u}$:

$$\begin{align*}
\text{minimize}_{u_0, \dot{u}} & \quad \| B \dot{u} \|_1 & \quad \text{(sparsity of coefficients)} \\
\text{subject to} & \quad \| D \dot{u} \|_2 \leq C & \quad \text{(smooth derivative)} \\
& \quad \left\| \begin{bmatrix} 1 & T \end{bmatrix} \begin{bmatrix} u_0 \\ \dot{u} \end{bmatrix} - \tilde{u} \right\|_2 \leq \gamma & \quad \text{(match a priori smoothed data).}
\end{align*}$$  \hspace{1cm} (11)

- Estimate $C$ using projection-based denoising result.

- Find $\gamma$ using the corner point of Pareto curve.
Learn coefficients and derivative simultaneously

- Write coefficients in terms of derivative:

\[
\Theta c = \dot{u} \quad \rightarrow \quad c = (\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T \dot{u}.
\]  

(10)

- Set \( B \) as estimator of \((\tilde{\Theta}^T \Theta)^{-1} \tilde{\Theta}^T\).

- Second order cone program (SOCP) to find initial condition, \( u_0 \), and derivative, \( \dot{u} \):

\[
\begin{align*}
\text{minimize}_{u_0, \dot{u}} & \quad \| B \dot{u} \|_1 & \quad \text{(sparsity of coefficients)} \\
\text{subject to} & \quad \| D \dot{u} \|_2 \leq C & \quad \text{(smooth derivative)} \\
& \quad \left\| \begin{bmatrix} 1 & T \end{bmatrix} \begin{bmatrix} u_0 \\ \dot{u} \end{bmatrix} - \tilde{u} \right\|_2 \leq \gamma & \quad \text{(match a priori smoothed data)}.
\end{align*}
\]  

(11)

- Estimate \( C \) using projection-based denoising result.

- Find \( \gamma \) using the corner point of Pareto curve.
SOCP improves derivative estimation compared with Tikhonov regularization
SOCP improves coefficient estimation compared with Lasso approach
Example prediction results: Van der Pol oscillator $u_1$
Example prediction results: Van der Pol oscillator $u_2$
Conclusions

- Existing equation discovery methods fail at high noise levels.
- Presented two improvements:
  1. Projection-based denoising strategy.
  2. SOCP to learn derivative/coefficients simultaneously.
- Led to improved derivative and coefficient estimation.
- Future Work:
  1. Compare approach to other versions of SINDy (i.e., Weak-Sindy).
  2. Consider these methods in the context of PDEs.
Conclusions

- Existing equation discovery methods fail at high noise levels.
- Presented two improvements:
  1. Projection-based denoising strategy.
  2. SOCP to learn derivative/coefficients simultaneously.
- Led to improved derivative and coefficient estimation.
- Future Work:
  1. Compare approach to other versions of SINDy (i.e, Weak-Sindy).
  2. Consider these methods in the context of PDEs.

Thank You!