MANCHESTER 1824

The University of Manchester

NLA Group Manchester

Inferring Chemical Reactions from Simulated Data

Department of Mathematics
The University of Manchester

ayana.mussabayeva@postgrad.manchester.ac.uk

Supervised by:
Stefan Güttel, Igor Larrosa

Contents

Catalytic Reactions

Catalytic reaction is a process of transformation of one or more reactants to products, accelerated or caused by substances called catalysts.

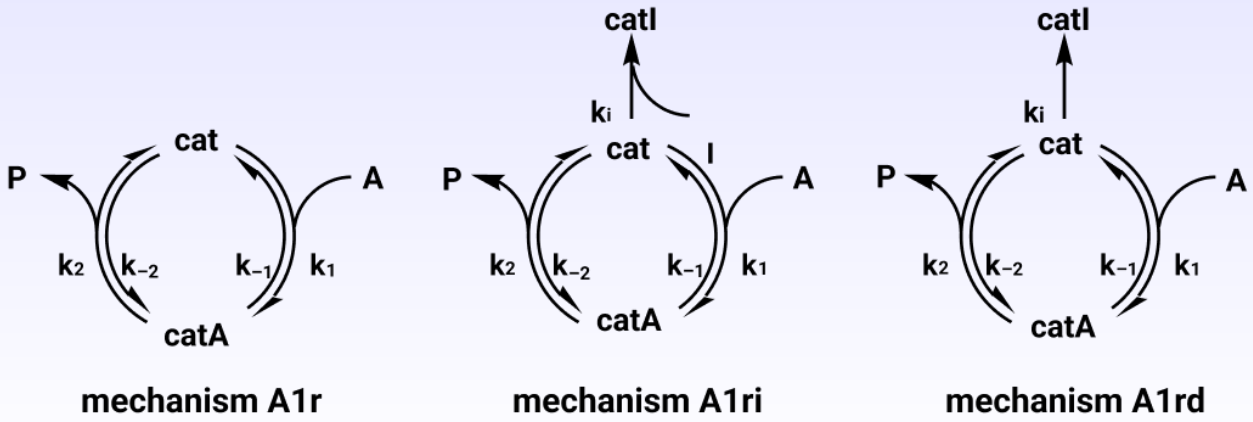
$$\mathrm{A} \xrightarrow{\mathrm{cat}} \mathrm{P}$$

Catalytic reactions are complex processes, and there may be different elementary reactions hiding under the above expression:

$$\mathrm{A} + \mathrm{cat} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{catA}$$

$$\mathrm{catA} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{P} + \mathrm{cat}$$

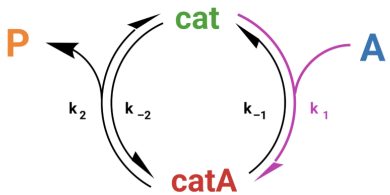
A **kinetic mechanism** shows the steps leading from the starting solutions to the formulation of intermediates and final products in chemical reactions. Kinetic mechanisms can be presented by diagrams (kinetic cycles):



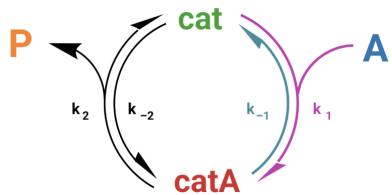
mechanism A1r | mechanism A1ri | mechanism A1rd

Kinetic Analysis

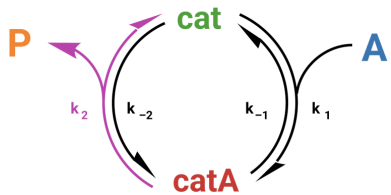
A kinetic mechanism can be represented as a system of ODEs:



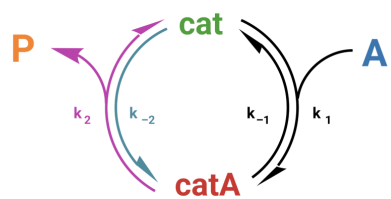
$$S_1 = -k_1[\mathrm{A}][\mathrm{cat}]$$



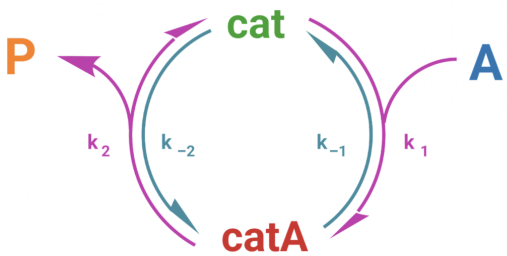
$$S_2 = +k_{-1}[\mathrm{catA}]$$



$$S_3 = +k_2[\mathrm{catA}]$$



$$S_4 = -k_{-2}[\mathrm{cat}][\mathrm{P}]$$



$$\frac{d[\mathrm{cat}]}{dt} = S_1 + S_2 + S_3 + S_4$$

$$\frac{d[\mathrm{cat}]}{dt} = -k_1[\mathrm{A}][\mathrm{cat}] + k_{-1}[\mathrm{catA}] + k_2[\mathrm{catA}] - k_{-2}[\mathrm{cat}][\mathrm{P}]$$

For *mechanism A1r* :

$$\begin{aligned}
[\dot{\mathrm{A}}] &= -k_1[\mathrm{A}][\mathrm{cat}] + k_{-1}[\mathrm{catA}] \\
[\dot{\mathrm{P}}] &= k_2[\mathrm{catA}] - k_{-2}[\mathrm{cat}][\mathrm{P}] \\
[\dot{\mathrm{cat}}] &= -k_1[\mathrm{A}][\mathrm{cat}] + k_2[\mathrm{catA}] + k_{-1}[\mathrm{catA}] - k_{-2}[\mathrm{catA}][\mathrm{P}] \\
[\dot{\mathrm{catA}}] &= k_1[\mathrm{A}][\mathrm{catA}] - k_2[\mathrm{catA}] - k_{-1}[\mathrm{catA}] + k_{-2}[\mathrm{catA}][\mathrm{P}]
\end{aligned}$$

Kinetic Analysis

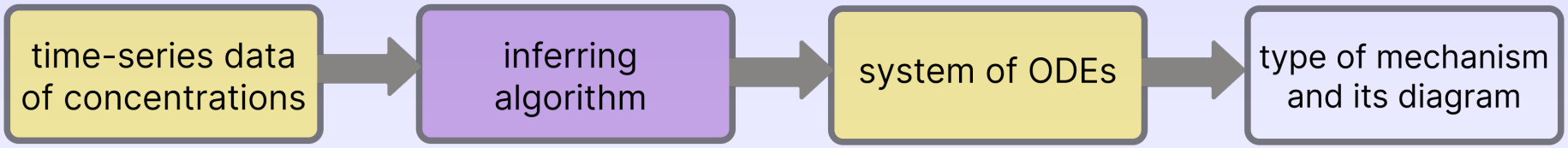
Kinetic mechanism shows the dependency of the reaction's rate with the kinetic constants and species' concentrations. Analyzing the kinetics of a reaction is essential for understanding the reaction itself and for improving its performance.

Classical methods for kinetic analysis in chemistry:

- Reaction progress kinetic analysis[1]
- Variable time normalization analysis[2]

Alternative methods:

- Neural networks[3]
- General-to-specific, simple-to-general iterative methods[4]



[1] Blackmond, "Reaction Progress Kinetic Analysis: A Powerful Methodology for Mechanistic Studies of Complex Catalytic Reactions".

[2] Burés, "Variable Time Normalization Analysis: General Graphical Elucidation of Reaction Orders from Concentration Profiles".

[3] Ji and Deng, "Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network".

[4] Srividhya et al., "Reconstructing biochemical pathways from time course data".

Mathematical Formulation

The input data concentrations are concatenated to form an $[k \times n]$ skinny matrix X, where k is the number of time points, and n is the number of species in the chemical reaction:

$$X = \begin{bmatrix} | & | & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & | & | \end{bmatrix}, \qquad X_{\text{A1r}} = \begin{bmatrix} | & | & | & | \\ [\text{A}] & [\text{P}] & [\text{cat}] & [\text{catA}] \\ | & | & | & | \end{bmatrix}.$$

The feature matrix $\widetilde{X}$ is a skinny matrix of size $[k \times m]$, where $m = \frac{n^2+3n}{2}$ representing the possible combinations of vectors of X, which can be found in the RHS of the ODE system:

$$\widetilde{X} = [X, \bar{X}, X^2] = \begin{bmatrix} | & \cdots & | & | & \cdots & | & | & \cdots & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n & \mathbf{x}_1\mathbf{x}_2 & \cdots & \mathbf{x}_{n-1}\mathbf{x}_n & \mathbf{x}_1^2 & \cdots & \mathbf{x}_n^2 \\ | & \cdots & | & | & \cdots & | & | & \cdots & | \end{bmatrix}.$$

The derivatives of the concentrations w.r.t. time form skinny matrix $\dot{X}$:

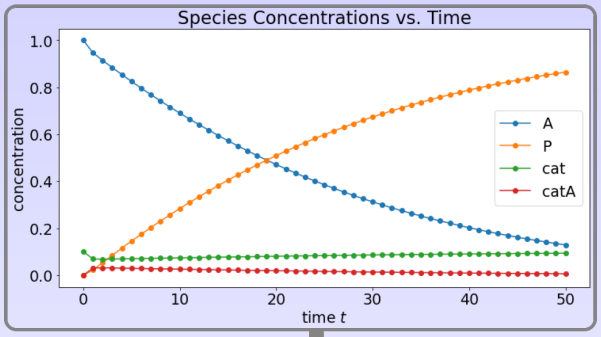
$$\dot{X} = \begin{bmatrix} | & | & | \\ F_1 & \cdots & F_n \\ | & | & | \end{bmatrix}, \qquad \dot{X}_{\text{A1r}} = \begin{bmatrix} | & | & | & | \\ [\dot{\text{A}}] & [\dot{\text{P}}] & [\dot{\text{cat}}] & [\dot{\text{catA}}] \\ | & | & | & | \end{bmatrix}.$$

Mathematical Formulation

The ODE system can be rewritten in a matrix form:

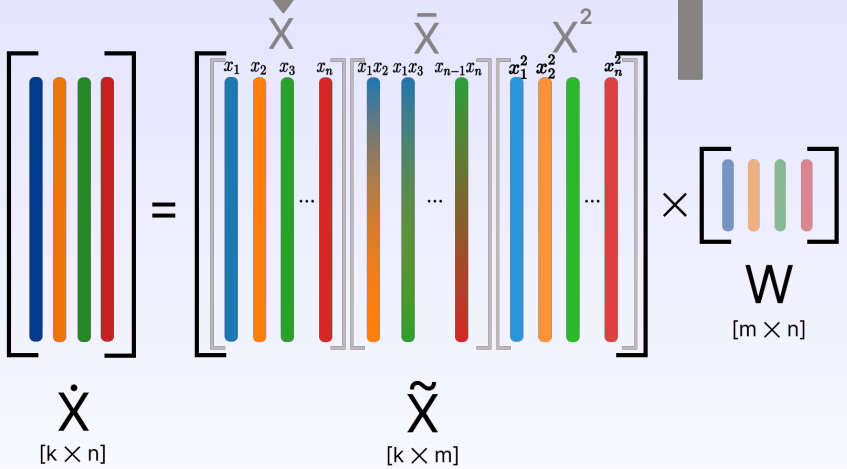
$$\dot{X} = \tilde{X} W,$$

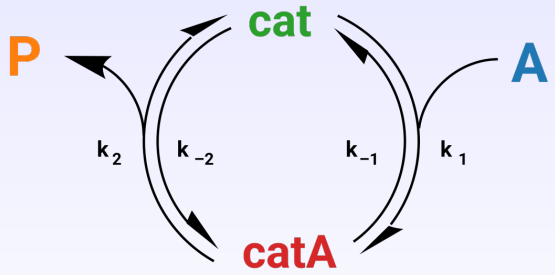
where W is a sparse matrix of coefficients of size $[m \times n]$.



system of ODEs

$$[\dot{\mathrm{A}}] = -k_1[\mathrm{A}][\mathrm{cat}] + k_{-1}[\mathrm{catA}]$$
$$[\dot{\mathrm{P}}] = k_2[\mathrm{catA}] - k_{-2}[\mathrm{cat}][\mathrm{P}]$$
$$[\dot{\mathrm{cat}}] = -k_1[\mathrm{A}][\mathrm{cat}] + k_2[\mathrm{catA}] + k_{-1}[\mathrm{catA}] - k_{-2}[\mathrm{cat}][\mathrm{P}]$$
$$[\dot{\mathrm{catA}}] = k_1[\mathrm{A}][\mathrm{cat}] - k_2[\mathrm{catA}] - k_{-1}[\mathrm{catA}] + k_{-2}[\mathrm{cat}][\mathrm{P}]$$



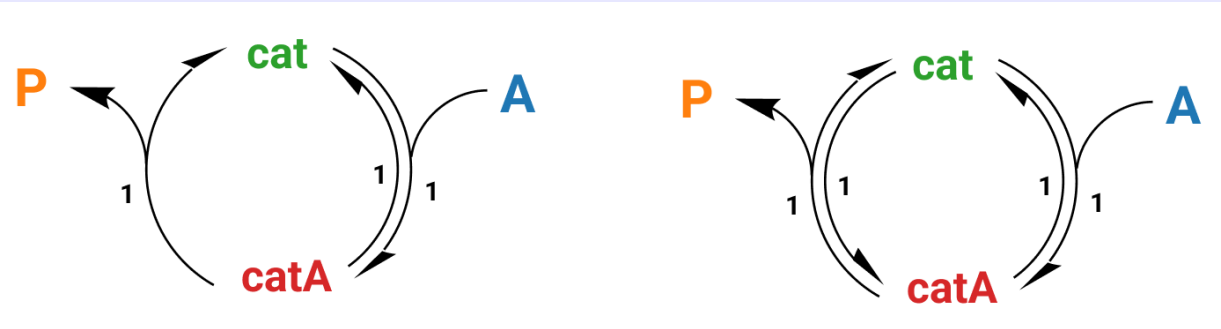


Data Simulation

The data is simulated by using the pre-known ODEs of **kinetic mechanism A1r**. Runge-Kutta-Felhberg 4(5) or **RKF45** method is used to approximate the solution of the pre-known ODEs.
We can use ℓ set of initial conditions for data generation. In that case ℓ matrices are stacked to form X, resulting the same size of $\dot{X}$:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_\ell \end{bmatrix}, \quad \dot{X} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \vdots \\ \dot{X}_\ell \end{bmatrix}.$$

- Kinetic constants: $k_1 = 1, k_{-1} = 1, k_2 = 1, k_{-2} = 0$
  $k_1 = 1, k_{-1} = 1, k_2 = 1, k_{-2} = 1$

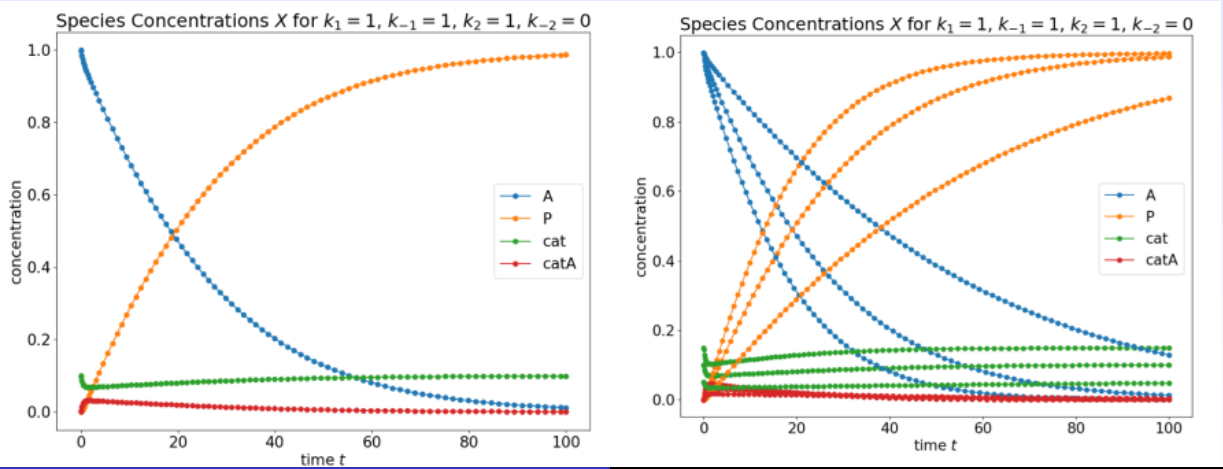


Data Simulation

The data is simulated by using the pre-known ODEs of **kinetic mechanism A1r**. Runge-Kutta-Felhberg 4(5) or **RKF45** method is used to approximate the solution of the pre-known ODEs.
We can use ℓ set of initial conditions for data generation. In that case ℓ matrices are stacked to form X, resulting the same size of $\dot{X}$:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_\ell \end{bmatrix}, \quad \dot{X} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \vdots \\ \dot{X}_\ell \end{bmatrix}.$$

- Kinetic constants: $k_1 = 1, k_{-1} = 1, k_2 = 1, k_{-2} = 0$
  $k_1 = 1, k_{-1} = 1, k_2 = 1, k_{-2} = 1$
- Initial conditions: $\ell = 3, \quad [\mathrm{A}]_0 = 1, [\mathrm{P}]_0 = 0, [\mathrm{catA}]_0 = 0,$
  $[\mathrm{cat}]_0 = \{0.05;\ 0.1;\ 0.15\}$



Derivatives Approximation

Matrix Y is introduced for numerical approximation of $\dot{X}$, which can be obtained using *finite difference approximation* or *univariate cubic spline approximation*[5]:

$$Y \approx \dot{X}.$$

The absolute and relative residual between the approximation and the ground truth vectors is found as :

$$\varepsilon_r(\dot{X}) = \frac{\|Y - \dot{X}\|_F}{\|\dot{X}\|_F} \qquad \varepsilon_{abs}(\dot{X}) = \|Y - \dot{X}\|_F$$

Optimizing the error in numerical differentiation can be achieved by *selecting appropriate time points*.
When using **equispaced points**:

	$[\text{cat}]_0 = 0.1$		$[\text{cat}]_0 = \{0.05, 0.1, 0.15\}$	
	$\varepsilon_{abs}(\dot{X})$	$\varepsilon_r(\dot{X})$	$\varepsilon_{abs}(\dot{X})$	$\varepsilon_r(\dot{X})$
Finite Difference	0.11356	0.53534	0.21324	0.56913
Cubic Spline	0.07880	0.038963	0.14838	0.41751

[5] Schultz, “Multivariate L-spline interpolation”.

Derivatives Approximation

Matrix Y is introduced for numerical approximation of $\dot{X}$, which can be obtained using *finite difference approximation* or *univariate cubic spline approximation*[5]:

$$Y \approx \dot{X}.$$

The absolute and relative residual between the approximation and the ground truth vectors is found as :

$$\varepsilon_r(\dot{X}) = \frac{\|Y - \dot{X}\|_F}{\|\dot{X}\|_F} \qquad \varepsilon_{abs}(\dot{X}) = \|Y - \dot{X}\|_F$$

Optimizing the error in numerical differentiation can be achieved by *selecting appropriate time points*.
When using the same **RKF45-chosen points**:

	$[\text{cat}]_0 = 0.1$		$[\text{cat}]_0 = \{0.05, 0.1, 0.15\}$	
	$\varepsilon_{abs}(\dot{X})$	$\varepsilon_r(\dot{X})$	$\varepsilon_{abs}(\dot{X})$	$\varepsilon_r(\dot{X})$
Finite Difference	0.06277	0.016409	0.11647	0.16406
Cubic Spline	0.00027	0.00078	0.00047	0.00073

[5] Schultz, “Multivariate L-spline interpolation”.

Derivatives Approximation

Matrix Y is introduced for numerical approximation of $\dot{X}$, which can be obtained using *finite difference approximation* or *univariate cubic spline approximation*[5]:

$$Y \approx \dot{X}.$$

The absolute and relative residual between the approximation and the ground truth vectors is found as :

$$\varepsilon_r(\dot{X}) = \frac{\|Y - \dot{X}\|_F}{\|\dot{X}\|_F} \qquad \varepsilon_{abs}(\dot{X}) = \|Y - \dot{X}\|_F$$

Optimizing the error in numerical differentiation can be achieved by *selecting appropriate time points*.
When using **Chebyshev points**:

	$[\text{cat}]_0 = 0.1$		$[\text{cat}]_0 = \{0.05, 0.1, 0.15\}$	
	$\varepsilon_{abs}(\dot{X})$	$\varepsilon_r(\dot{X})$	$\varepsilon_{abs}(\dot{X})$	$\varepsilon_r(\dot{X})$
Finite Difference	0.05936	0.15783	0.11144	0.16060
Cubic Spline	0.00013	0.00038	0.00025	0.00039

[5] Schultz, “Multivariate L-spline interpolation”.

Performance Metrics

Using the inferred weights W, we can compute the relative distance of the inferred derivatives $\widehat{Y} = \widetilde{X}W$ to the spline approximation Y and the distance of the forward-solved $\widehat{X}$ to the data X:

$$\varepsilon'_r(\widehat{Y}) = \frac{\|\widehat{Y} - Y\|_F}{\|Y\|_F}, \qquad \varepsilon_r(\widehat{X}) = \frac{\|\widehat{X} - X\|_F}{\|X\|_F}.$$

By knowing the exact system of ODEs for mechanism A1r, which was used for generating X, we also calculate:

$$\varepsilon_r(\widehat{Y}) = \frac{\|\widehat{Y} - \dot{X}\|_F}{\|\dot{X}\|_F}, \qquad \varepsilon_r(W) = \frac{\|W - W_{\text{ex}}\|_F}{\|W_{\text{ex}}\|_F},$$

where W_{ex} is a $[4 \times 14]$ s-sparse matrix:

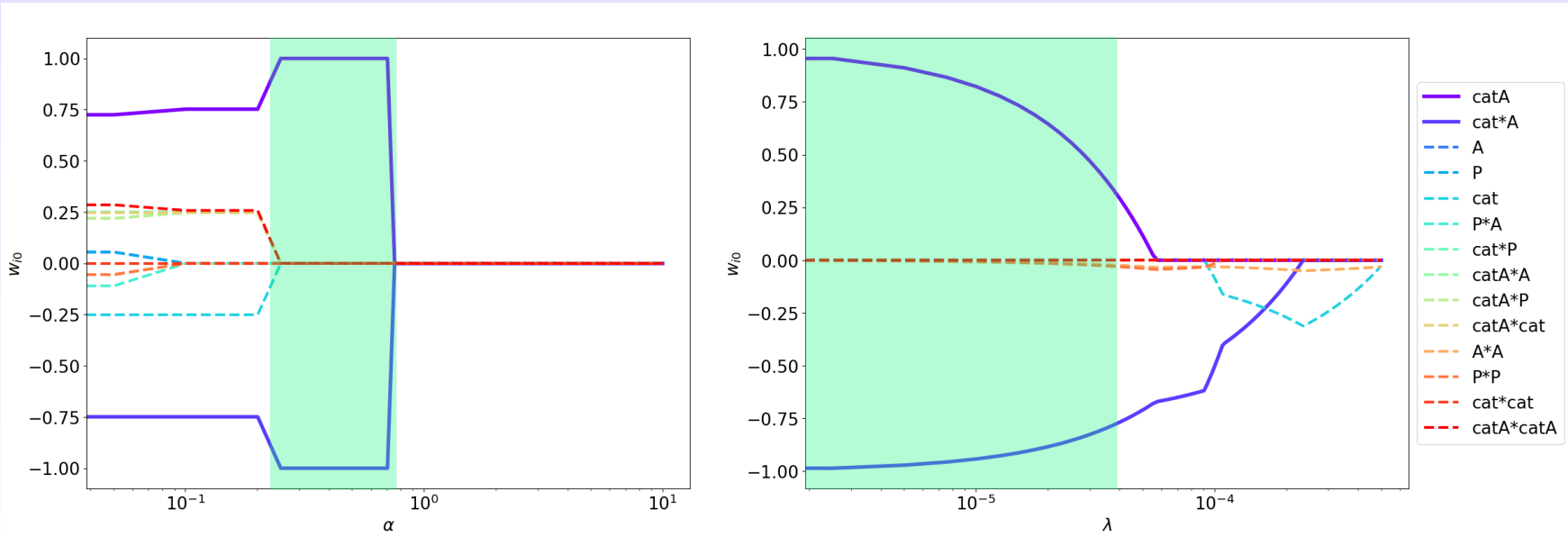
$$W_{\text{ex}} = \begin{bmatrix} 0 & 0 & 0 & k_{-1} & 0 & -k_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & k_2 & 0 & 0 & -k_{-2} & 0 & \dots & 0 \\ 0 & 0 & 0 & k_2 + k_{-1} & 0 & -k_1 & -k_{-2} & 0 & \dots & 0 \\ 0 & 0 & 0 & -k_2 - k_{-1} & 0 & k_1 & k_{-2} & 0 & \dots & 0 \end{bmatrix}.$$

Performance Metrics

Regularization paths show the coefficients of the features $\widetilde{X}$ for different values of some optimization parameter(either thresholding (α) or regularization (λ)) parameters for each equation.

The solid lines represent the coefficients which should be non-zero, or the coefficients of the "active" components. The dashed lines show the coefficients of those components, which should have zero coefficients.

$$[\dot{\mathrm{A}}] = -k_1[\mathrm{A}][\mathrm{cat}] + k_{-1}[\mathrm{catA}]$$



Sparse Identification of Non-linear Dynamics

STLSQ is used in **pysindy**[6] library for sparse identification of non-linear dynamics (SINDy)[7]. We have tried using SINDy model with thresholding and Ridge regularization. SINDy models fail to identify the correct active components and the implementation is not very flexible. The best result obtained by SINDy with thresholding $\alpha = 0.01$ and Ridge regularization $\lambda = 0.05$:

$$\begin{aligned}
\dot{[\mathrm{A}]} &= 1.01[\mathrm{catA}] - 1.00[\mathrm{A}][\mathrm{cat}] \\
\dot{[\mathrm{P}]} &= 1.01[\mathrm{catA}] - 0.01[\mathrm{catA}][\mathrm{cat}] - 0.01[\mathrm{A}][\mathrm{A}] \\
\dot{[\mathrm{cat}]} &= 2.03[\mathrm{catA}] - 1.01[\mathrm{catA}][\mathrm{cat}] + 0.02[\mathrm{A}][\mathrm{A}] \\
\dot{[\mathrm{catA}]} &= -2.03[\mathrm{catA}] + 1.01[\mathrm{catA}][\mathrm{cat}] + 0.02[\mathrm{A}][\mathrm{A}].
\end{aligned}$$

While the desired system of ODEs:

$$\begin{aligned}
\dot{[\mathrm{A}]} &= [\mathrm{catA}] - [\mathrm{A}][\mathrm{cat}] \\
\dot{[\mathrm{P}]} &= [\mathrm{catA}] \\
\dot{[\mathrm{cat}]} &= 2[\mathrm{catA}] - [\mathrm{A}][\mathrm{cat}] \\
\dot{[\mathrm{catA}]} &= -2[\mathrm{catA}] + [\mathrm{A}][\mathrm{cat}].
\end{aligned}$$

[6] Silva et al., "PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data".

[7] Brunton, Proctor, and Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems".

Unregularized Least Squares Solution

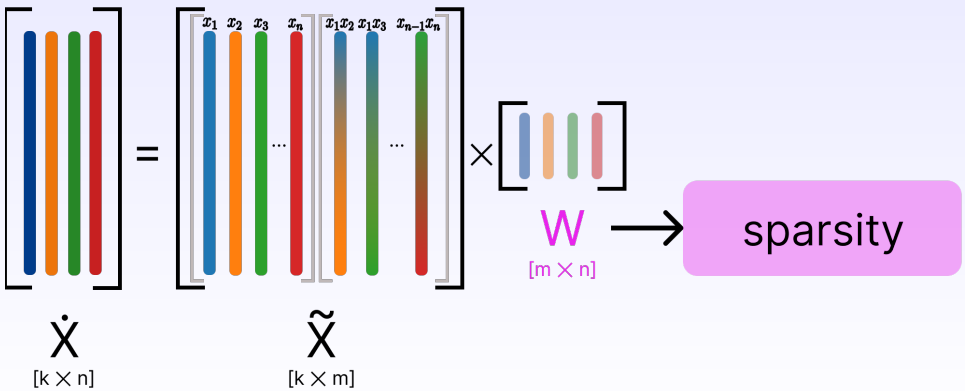
We need to find a sparse matrix of coefficients W, such that:

$$Y \approx \widetilde{X} W.$$

Minimizing the second norm of the difference between the column y_j and $\widetilde{X} w_j$ corresponding to each ODE:

$$w_j = \underset{\hat{w}}{\operatorname{argmin}} \|y_j - \widetilde{X}\hat{w}_j\|_2.$$

Sampling	Approximation	$\varepsilon'_r(\widehat{Y})$	$\varepsilon_r(\widehat{Y})$	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
RKF45	finite difference	0.2774	0.1025	0.0128	0.9933	56
RKF45	spline	0.0008	0.0001	0.000002	1.0297	56
Equispaced	finite difference	0.0474	0.0229	0.0107	0.9995	56
Equispaced	spline	0.3216	0.0171	0.0043	0.9978	56
Chebyshev	finite difference	0.2641	0.0926	0.0140	0.9932	56
Equispaced	spline	0.0004	0.0001	0.000003	1.0241	56



- Lasso regularization
- Thresholding

Sequentially Thresholded Least Squares

Sequentially thresholded least squares (STLSQ) iteratively computes least squares solutions

$$y_j \approx \widetilde{X} w_j, \qquad j = 1, \ldots, n,$$

and zeroes entries w_{ij} of w_j if $|w_{ij}| < \alpha$.

Sampling	Approximation	α	$\varepsilon'_r(\widehat{Y})$	$\varepsilon_r(\widehat{Y})$	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
Equispaced	finite difference	$\alpha = 0.4$	0.4762	0.0267	0.0149	0.9995	56
Equispaced	spline	$\alpha = 0.4$	0.3224	0.0163	0.0002	0.9978	56
RKF45	finite difference	$\alpha = 0.4$	0.2774	0.1025	0.0128	0.9932	54
RKF45	spline	$\alpha = 0.5$	0.0368	0.0368	1.4108	0.7562	5
RKF45	spline	$\alpha = 0.4$	0.0008	0.0001	**0.000008**	**0.0003**	**7**
Chebyshev	finite difference	$\alpha = 0.4$	1.0269	0.8246	0.1891	0.9931	54
Chebyshev	spline	$\alpha = 0.4$	0.7986	0.7984	0.1814	**0.00004**	**7**

For RKF45-chosen points with spline approximation and $\alpha = 0.4$, the obtained ODEs are:

$$\begin{aligned}
\dot{[\mathrm{A}]} &= 0.999568[\mathrm{catA}] - 0.999759[\mathrm{cat}][\mathrm{A}] \\
\dot{[\mathrm{P}]} &= 1.00004[\mathrm{catA}] \\
\dot{[\mathrm{cat}]} &= 1.999384[\mathrm{catA}] - 0.999652[\mathrm{cat}][\mathrm{A}] \\
\dot{[\mathrm{catA}]} &= -1.999384[\mathrm{catA}] + 0.999652[\mathrm{cat}][\mathrm{A}].
\end{aligned}$$

Sequentially Thresholded Least Squares

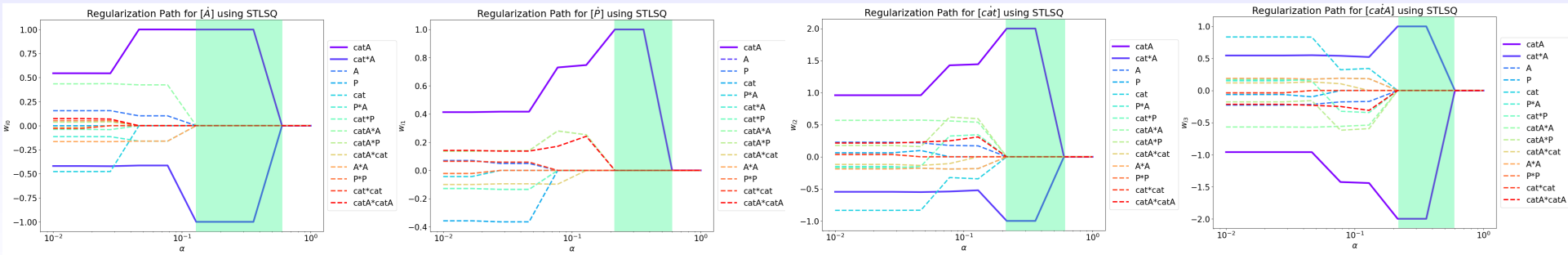
Sequentially thresholded least squares (STLSQ) iteratively computes least squares solutions

$$y_j \approx \widetilde{X} w_j, \qquad j = 1, \ldots, n,$$

and zeroes entries w_{ij} of w_j if $|w_{ij}| < \alpha$.

Sampling	Approximation	α	$\varepsilon'_r(\widehat{Y})$	$\varepsilon_r(\widehat{Y})$	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
Equispaced	finite difference	$\alpha = 0.4$	0.4762	0.0267	0.0149	0.9995	56
Equispaced	spline	$\alpha = 0.4$	0.3224	0.0163	0.0002	0.9978	56
RKF45	finite difference	$\alpha = 0.4$	0.2774	0.1025	0.0128	0.9932	54
RKF45	spline	$\alpha = 0.5$	0.0368	0.0368	1.4108	0.7562	5
RKF45	spline	$\alpha = 0.4$	0.0008	0.0001	**0.000008**	**0.0003**	**7**
Chebyshev	finite difference	$\alpha = 0.4$	1.0269	0.8246	0.1891	0.9931	54
Chebyshev	spline	$\alpha = 0.4$	0.7986	0.7984	0.1814	**0.00004**	**7**

Regularization paths for RKF45-chosen points with spline approximation, $\alpha = 0.4$:



Sequentially Thresholded Least Squares

Sequentially thresholded least squares (STLSQ) iteratively computes least squares solutions

$$y_j \approx \widetilde{X} w_j, \qquad j = 1, \ldots, n,$$

and zeroes entries w_{ij} of w_j if $|w_{ij}| < \alpha$.

Sampling	Approximation	α	$\varepsilon'_r(\widehat{Y})$	$\varepsilon_r(\widehat{Y})$	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
Equispaced	finite difference	$\alpha = 0.4$	0.4762	0.0267	0.0149	0.9995	56
Equispaced	spline	$\alpha = 0.4$	0.3224	0.0163	0.0002	0.9978	56
RKF45	finite difference	$\alpha = 0.4$	0.2774	0.1025	0.0128	0.9932	54
RKF45	spline	$\alpha = 0.5$	0.0368	0.0368	1.4108	0.7562	5
RKF45	spline	$\alpha = 0.4$	0.0008	0.0001	**0.000008**	**0.0003**	**7**
Chebyshev	finite difference	$\alpha = 0.4$	1.0269	0.8246	0.1891	0.9931	54
Chebyshev	spline	$\alpha = 0.4$	0.7986	0.7984	0.1814	**0.00004**	**7**

For Chebyshev points with spline approximation and $\alpha = 0.4$, the obtained ODEs are:

$$\begin{aligned}
[\dot{\mathrm{A}}] &= 1.000032[\mathrm{catA}] - 1.000002[\mathrm{cat}][\mathrm{A}] \\
[\dot{\mathrm{P}}] &= 1.000013[\mathrm{catA}] \\
[\dot{\mathrm{cat}}] &= 2.000051[\mathrm{catA}] - 1.000005[\mathrm{cat}][\mathrm{A}] \\
[\dot{\mathrm{catA}}] &= -2.0000515[\mathrm{catA}] + 1.000005[\mathrm{cat}][\mathrm{A}].
\end{aligned}$$

Sequentially Thresholded Least Squares

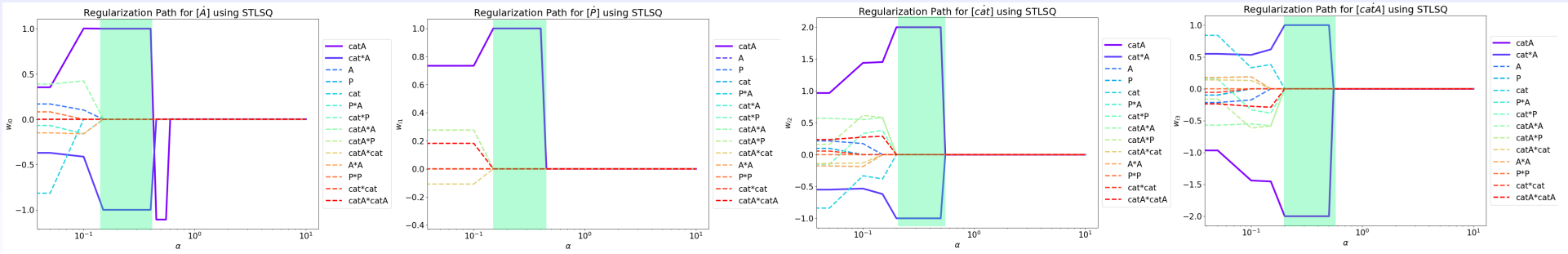
Sequentially thresholded least squares (STLSQ) iteratively computes least squares solutions

$$y_j \approx \widetilde{X} w_j, \qquad j = 1, \ldots, n,$$

and zeroes entries w_{ij} of w_j if $|w_{ij}| < \alpha$.

Sampling	Approximation	α	$\varepsilon_r'(\widehat{Y})$	$\varepsilon_r(\widehat{Y})$	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
Equispaced	finite difference	$\alpha = 0.4$	0.4762	0.0267	0.0149	0.9995	56
Equispaced	spline	$\alpha = 0.4$	0.3224	0.0163	0.0002	0.9978	56
RKF45	finite difference	$\alpha = 0.4$	0.2774	0.1025	0.0128	0.9932	54
RKF45	spline	$\alpha = 0.5$	0.0368	0.0368	1.4108	0.7562	5
RKF45	spline	$\alpha = 0.4$	0.0008	0.0001	**0.000008**	**0.0003**	**7**
Chebyshev	finite difference	$\alpha = 0.4$	1.0269	0.8246	0.1891	0.9931	54
Chebyshev	spline	$\alpha = 0.4$	0.7986	0.7984	0.1814	**0.00004**	**7**

Regularization paths for Chebyshev points with spline approximation, $\alpha = 0.4$:



Lasso Regularization

Lasso regularization promotes sparsity by adding an ℓ_1-norm regularization term involving the coefficient vector with parameter λ:

$$w_j = \underset{\hat{w}}{\operatorname{argmin}} \|y_j - \widetilde{X}\hat{w}\|_2^2 + \lambda\|\hat{w}\|_1.$$

Sampling	Approximation	λ	$\varepsilon_r'(\widehat{Y})$	$\varepsilon_r(\widehat{Y})$	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
Equispaced	finite difference	λ = 0.001	0.4307	0.2757	0.0229	3.064	10
Equispaced	finite difference	λ = 1e-05	0.9997	0.9997	1.2282	1.1272	20
Equispaced	spline	λ = 0.001	0.9105	0.8904	0.0508	7.5710	11
Equispaced	spline	λ = 1e-05	0.3219	0.0896	0.0025	1.9061	14
RKF45	finite difference	λ = 0.001	1.0732	0.8596	0.2589	0.7275	14
RKF45	finite difference	λ = 1e-05	1.0275	0.8181	0.2618	0.9990	16
RKF45	spline	λ = 0.001	0.8014	0.8011	0.2502	0.1486	7
RKF45	spline	λ = 1e-05	0.7800	0.7798	**0.00005**	**0.0016**	**7**
Chebyshev	finite difference	λ = 0.001	0.2469	0.1211	0.0104	1.2316	13
Chebyshev	finite difference	λ = 1e-5	0.2469	0.1211	0.0104	1.2316	13
Chebyshev	finite difference	λ = 0.01	0.0366	0.0367	0.0076	0.1489	8
Chebyshev	spline	λ = 1e-5	0.0365	0.0367	**0.00006**	**0.0013**	**7**

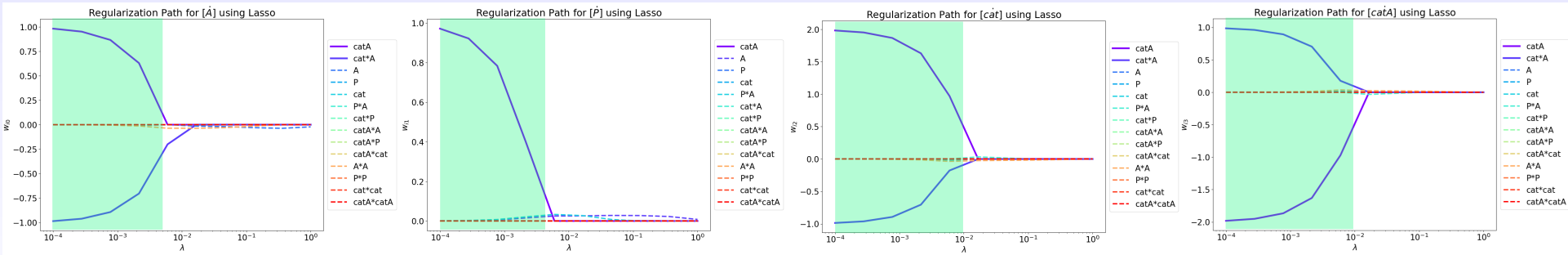
Lasso Regularization

Lasso regularization promotes sparsity by adding an ℓ_1-norm regularization term involving the coefficient vector with parameter λ:

$$w_j = \underset{\hat{w}}{\operatorname{argmin}} \|y_j - \widetilde{X}\hat{w}\|_2^2 + \lambda\|\hat{w}\|_1.$$

When using RKF45-chosen points and Lasso regularization with $\lambda = 10^{-5}$:

$$\begin{aligned}
[\dot{\mathrm{A}}] &= 1.00[\mathrm{catA}] - 1.00[\mathrm{cat}][\mathrm{A}] \\
[\dot{\mathrm{P}}] &= 1.00[\mathrm{catA}] \\
[\dot{\mathrm{cat}}] &= 2.00[\mathrm{catA}] - 1.00[\mathrm{cat}][\mathrm{A}] \\
[\dot{\mathrm{catA}}] &= -2.00[\mathrm{catA}] + 1.00[\mathrm{cat}][\mathrm{A}].
\end{aligned}$$



When using Chebyshev sampling, the result is approximately the same

Concluding Remarks

In order to infer kinetics from catalytic reaction's data it is necessary to consider the *sampling of solution; numerical differentiation techniques; regularization methods*.

Solver	Sampling	Approximation	$\varepsilon_r(\widehat{X})$	$\varepsilon_r(W)$	s
Unregularized	Equispaced	finite difference	0.0107	0.9995	56
Unregularized	Equispaced	spline	0.0043	0.9978	56
Unregularized	RKF45	finite difference	0.0128	0.9933	56
Unregularized	RKF45	spline	0.000002	1.0297	56
Unregularized	Chebyshev	finite difference	0.0140	0.9932	56
Unregularized	Chebyshev	spline	0.000003	1.0241	56
STLSQ, α = 0.4	Equispaced	finite difference	0.0149	0.9995	56
STLSQ, α = 0.4	Equispaced	spline	0.0002	0.9978	56
STLSQ, α = 0.4	RKF45	finite difference	0.0128	0.9932	54
STLSQ, α = 0.4	RKF45	spline	**0.000008**	**0.0003**	**7**
STLSQ, α = 0.4	Chebyshev	finite difference	0.1891	0.9931	54
STLSQ, α = 0.4	Chebyshev	spline	0.1814	**0.00004**	**7**
Lasso, λ = 1e-5	Equispaced	finite difference	1.2282	1.1272	20
Lasso, λ = 1e-5	Equispaced	spline	0.0025	1.9061	14
Lasso, λ = 1e-5	RKF45	finite difference	0.2618	0.9990	16
Lasso, λ = 1e-5	RKF45	spline	**0.00005**	**0.0016**	**7**
Lasso, λ = 1e-5	Chebyshev	finite difference	0.0076	0.1489	8
Lasso, λ = 1e-5	Chebyshev	spline	**0.00006**	**0.0013**	**7**

Concluding Remarks

Summary:

- It is crucial to minimize numerical error for inferring kinetic mechanism
- Both Lasso and STLSQ can provide accurate result
- Lasso is more resistant to numerical error than STLSQ

The limitations and future work:

- Simulated non-noisy data should be replaced with experimental data
- The methods need to be tested using more complex kinetic mechanisms

MANCHESTER 1824
The University of Manchester

Thank you for your attention!