

Local vs Global Approaches to Model Learning: Algorithms, Failures and Theorem

François Fages

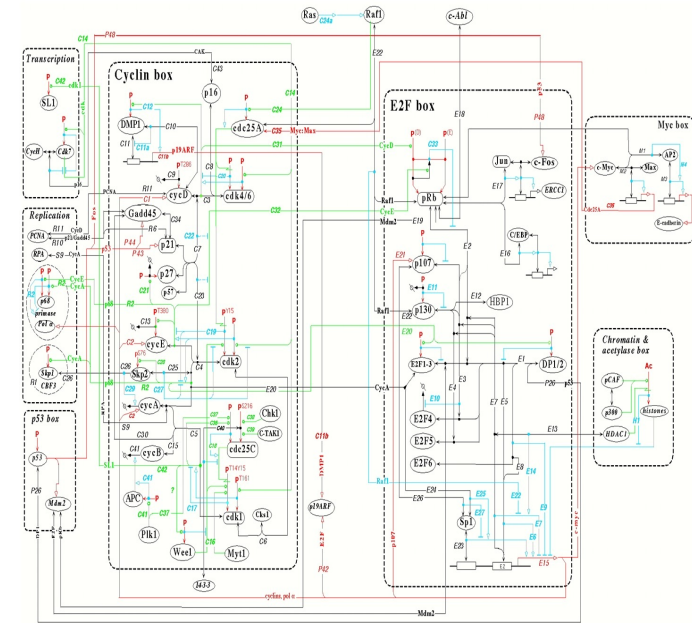
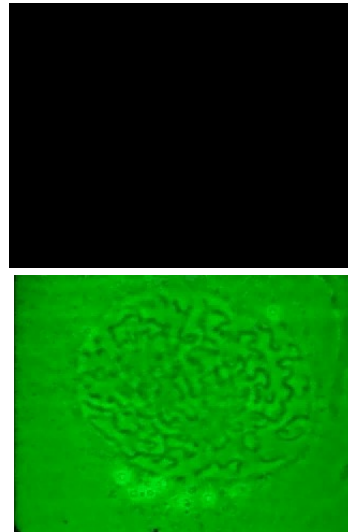
joint work with J. Martinelli, A. El Gadarri, A. Tan Lhrnould and others

Lifeware group <http://lifeware.inria.fr/>
on Computational Systems Biology and Optimization

Inria Saclay
Palaiseau, France

Motivation for this Work: Computational Systems Biology

- Cells process signals
- Take decisions such as
 - Change of metabolism
 - Cell differentiation
 - Replication, cell division
 - Moving
 - Apoptosis
- Control the execution of those processes



Programs? Information encoded in molecular concentrations, Chemical Reaction Networks (CRNs) [Feinberg 1977]

- Analog computation with proteins: continuous concentration levels, continuous time, CRN interpreted by ODE
- Analog programs defined by (digital) gene expression through the presence or not of gene encoded proteins

Systems Biology: model cell processes with CRNs, understand natural CRN programs and their function

Synthetic Biology: design CRNs to implement particular functions (in reactors, artificial vesicles or cells).

Turing completeness theorem [F., Le Guludec, Bournez, Pouly, CMSB 2017] *Any computable real function can be computed by an elementary CRN (mass action law reactions with at most 2 reactants) with ODE interpretation.*

Outline of my Talk

1. Chemical Reaction Networks as a programming language
 - Syntax: [reactions with rate functions](#)
 - Hierarchy of semantics: [continuous ODE](#), stochastic CTMC, Petri Net, asynchronous Boolean transition system
2. Learning protocol for inferring CRNs from time series data
 - Example of videomicroscopy data
 - Simulation traces obtained from a hidden CRN
3. Global approach: SINDy non-linear regression algorithm for learning ODE models
 - Failures on simple simulation examples
4. Local approach: Reactmine stochastic search algorithm for learning CRN models
 - Results on simulation examples
5. Sufficient conditions for the correctness of SINDy Lasso regression
6. Conclusion and on-going work

1. CRN Syntax

Let $S = \{x_1, \dots, x_s\}$ be a finite set of molecular species.

Def. A reaction is a quadruple (R, I, P, f) , also noted $R / I \xrightarrow{f} P$

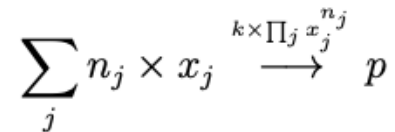
where R (resp. I, P) is a multiset of reactant species (resp. inhibitors, products)

and $f: \mathbb{R}_+^S \rightarrow \mathbb{R}_+$ is a rate function (kinetic expression).

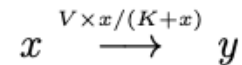
- Multisets are represented by linear expressions with integer stoichiometric coefficients
- A reaction catalyst is a molecular species that is both a reactant and a product (can also be an inhibitor).

E.g. reactions with

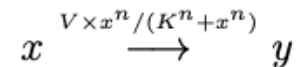
- Mass action law kinetics



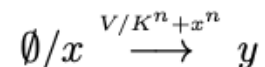
- Michaelis-Menten kinetics



- Hill kinetics



- Negative Hill kinetics



Well-formed Reactions

Def. A reaction (R, I, P, f) is **well-formed** if

- $f: \mathbb{R}_+^S \rightarrow \mathbb{R}_+$ is a partially differentiable function
- $x_i \in R$ if and only if $\frac{\partial f}{\partial x_i}(x) > 0$ for some value $x \in \mathbb{R}_+^S$
- $x_i \in I$ if and only if $\frac{\partial f}{\partial x_i}(x) < 0$ for some value $x \in \mathbb{R}_+^S$.

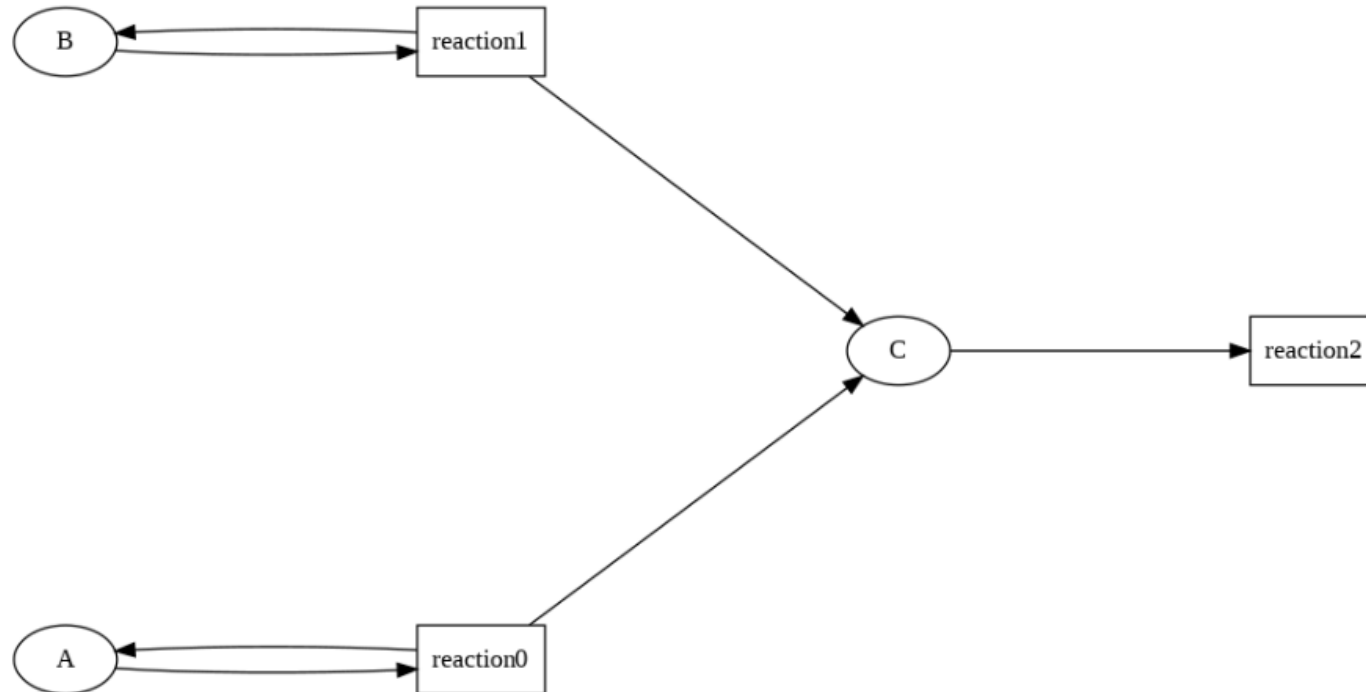
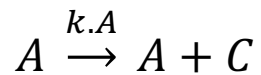
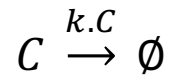
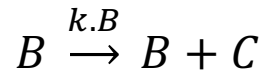
Def. A reaction is **strict** if $R(x_i) > 0$ implies $f(x_1, \dots, x_s) = 0$ whenever $x_i = 0$.

Prop. The ODE associated to a well-formed and strict reaction system (CRN) defines a positive system.

F., Gay, Soliman. [Inferring Reaction Systems from Ordinary Differential Equations](#). *Theoretical Computer Science*, 599:64–78, 2015.

Hypergraph Structure of a Reaction System

Standard representation of a hypergraph by a [bipartite species/reaction graph](#).



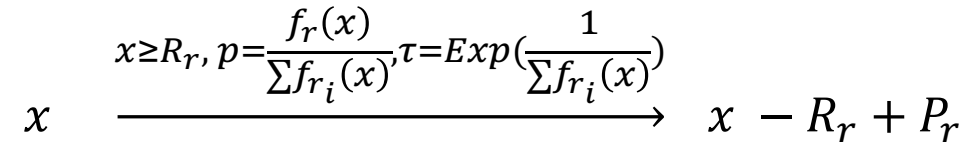
CRN Semantics

One given reaction system $\{(R_r, I_r, P_r, f_r)\}_{r \in \mathcal{C}}$ can be interpreted in different formalisms:

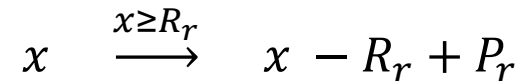
- Continuous interpretation by **ordinary differential equations (ODE)** on $x \in \mathbb{R}_+^S$

$$\frac{dx_i}{dt} = \sum_{r \in \text{CRN}} (P_r(x_i) - R_r(x_i)) \cdot f_r(x)$$

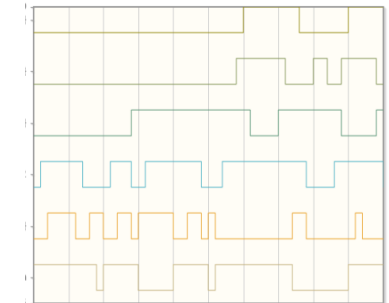
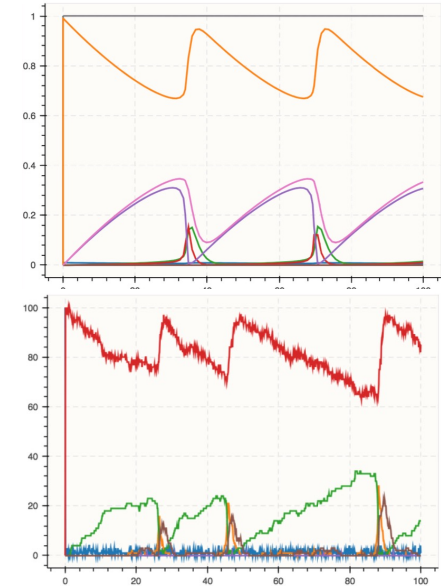
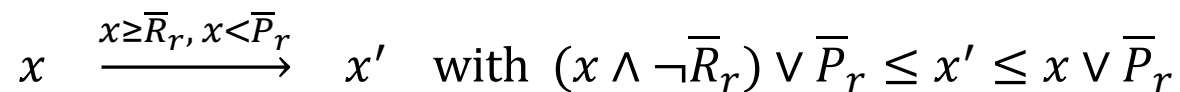
- Stochastic interpretation by **continuous-time Markov chain (CTMC)** $x \in \mathbb{N}_+^S$



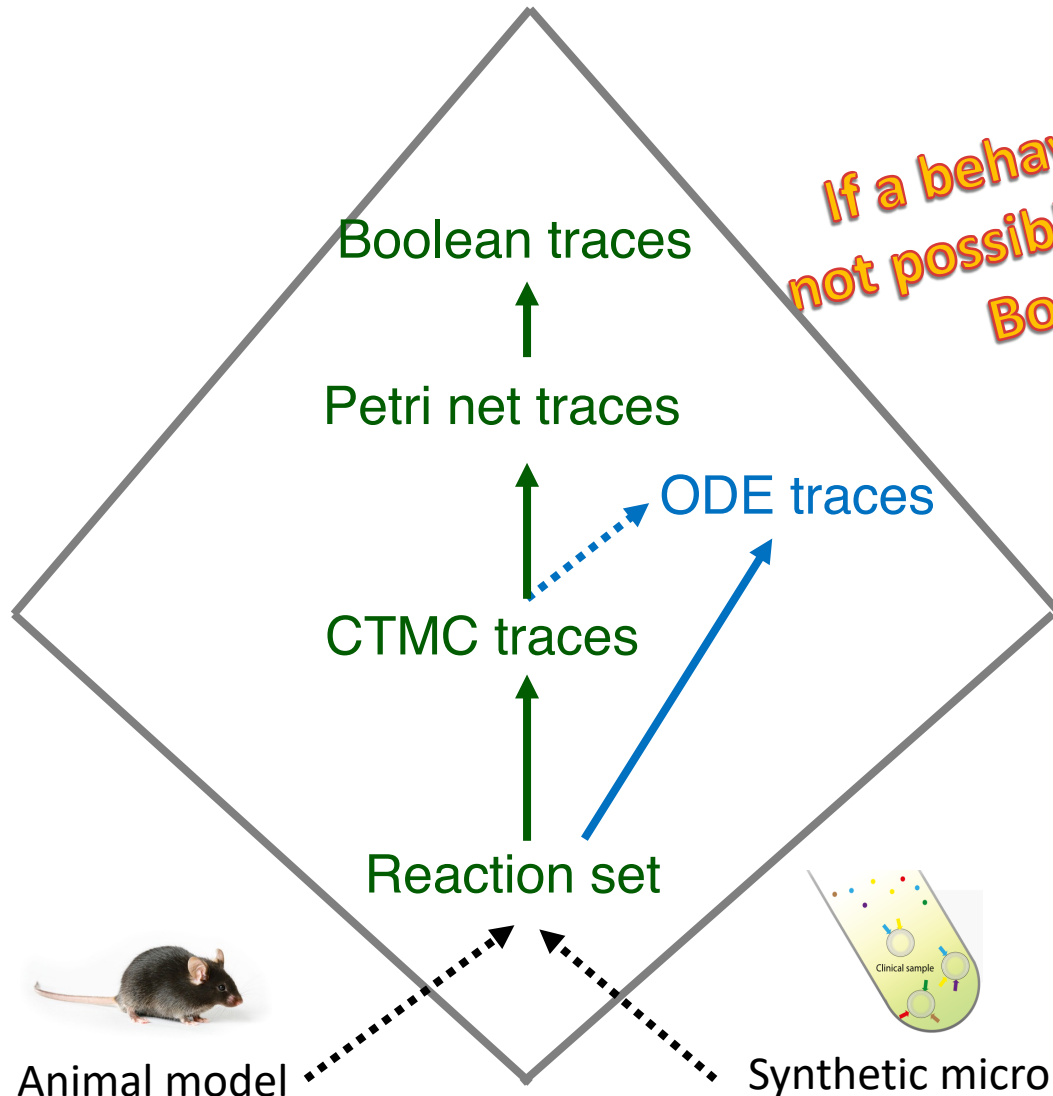
- Rate-independent non-deterministic discrete interpretation by **Petri Net (PN)** $x \in \mathbb{N}_+^S$



- Rate-independent **asynchronous Boolean state transition** interpretation $x \in \mathcal{B}^S$



Hierarchy of Semantics



If a behavior is not possible in the Boolean semantics, it is not possible in the stochastic semantics for any reaction rates. Boolean model-checking can exhibit rare events

Thm. (abstract interpretation \uparrow) Galois connections between the domains of syntactical, stochastic, Petri net and Boolean trace semantics

Fages, Soliman. [Abstract Interpretation and Types for Systems Biology](#).

Theoretical Computer Science, 403(1):52–70, 2008.

Thm. (approximation \dashrightarrow) [Gillespie 1971 Kurtz 1978, 1992] When the volume tends to infinity the ODE trace approximates the *mean* stochastic trace

Thm. (equality) [Buscemi Fages CMSB 2024] Under graphical conditions on the ancestors of polymolecular reactions, the ODE trajectory equals the *mean* stochastic trace.

Compiling Cosine(time) in BIOCHAM



`biocham`: `compile_from_expression(cos,time,f)`.

`initial_state(f_p=1)`.

MA(1.0) for $f_p \Rightarrow A_m + f_p$.

MA(1.0) for $f_m \Rightarrow A_p + f_m$.

MA(1.0) for $A_p \Rightarrow A_p + f_p$.

MA(1.0) for $A_m \Rightarrow A_m + f_m$.

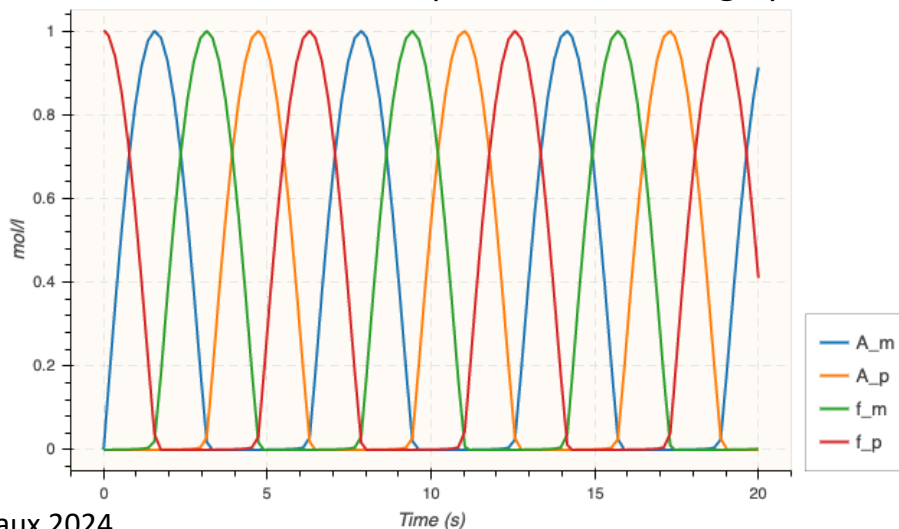
MA(fast) for $f_m + f_p \Rightarrow _$.

MA(fast) for $A_m + A_p \Rightarrow _$.

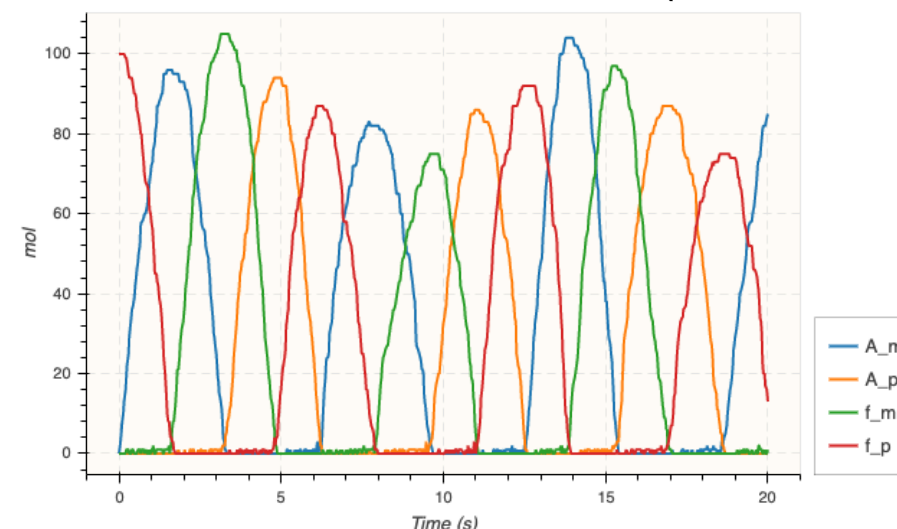
ODE:

$$\begin{aligned}\frac{dA_m}{dt} &= f_p - fast * A_m * A_p \\ \frac{dA_p}{dt} &= f_m - fast * A_m * A_p \\ \frac{df_m}{dt} &= A_m - fast * f_m * f_p \\ \frac{df_p}{dt} &= A_p - fast * f_m * f_p\end{aligned}$$

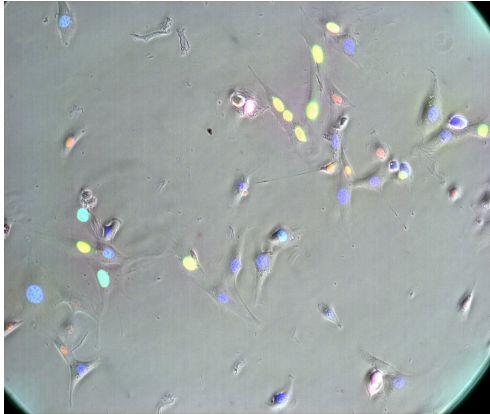
ODE simulation (for CRN design)



Stochastic CTMC simulation (for CRN testing)

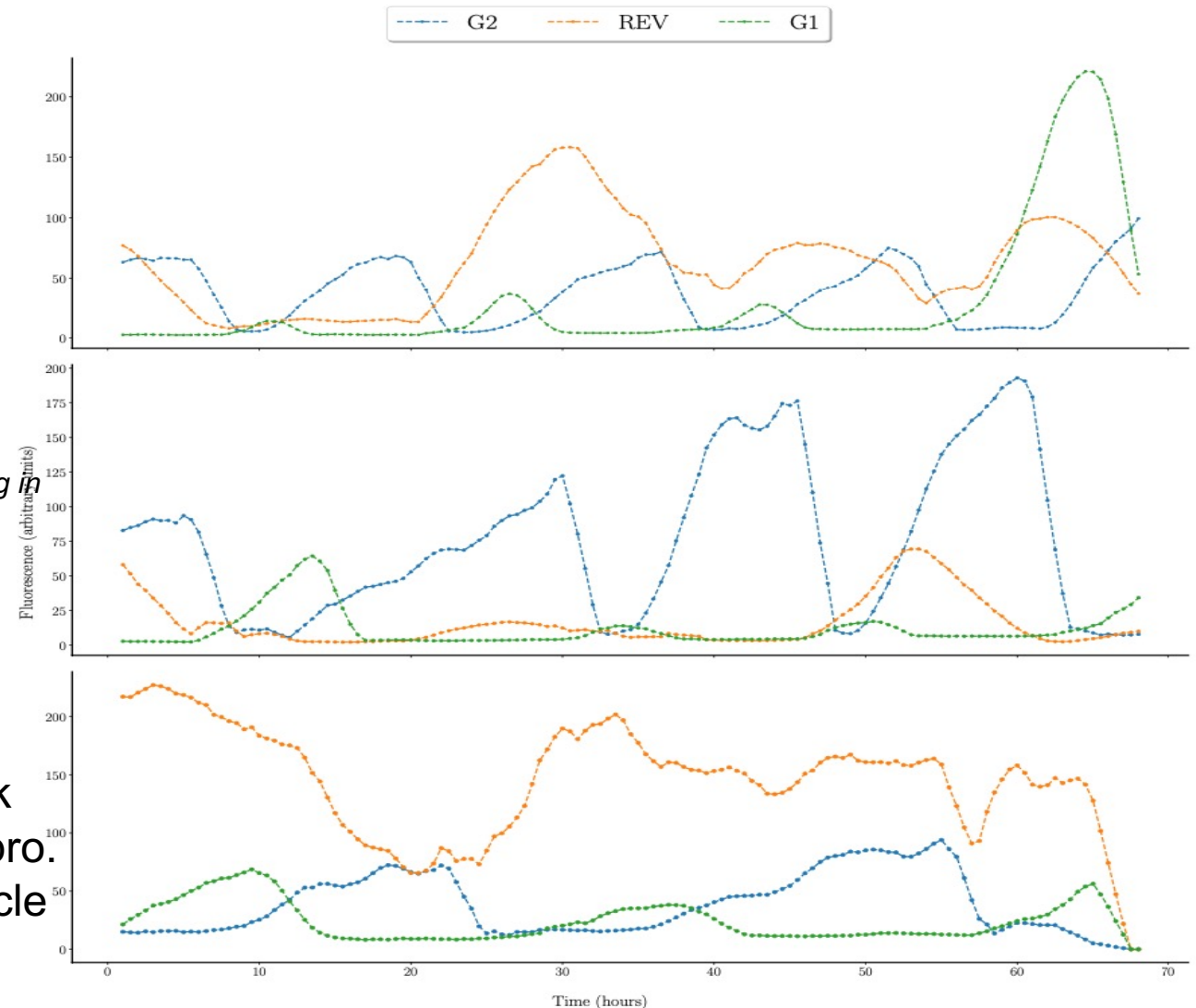


2. Single Cell Videomicroscopy Data (67 cells during 48h)



Pauline Traynard, Céline Feillet, Sylvain Soliman, Franck Delaunay, F-Model-based Investigation of the Circadian Clock and Cell Cycle Coupling in Mouse Embryonic Fibroblasts: Prediction of RevErb-alpha Up-Regulation during Mitosis. *Biosystem*, 149:59–69, 2016.

- Markers of cell cycle and circadian clock
 - Single “wild type” trace of single cells
 - Noisy data, statistical analysis, clustering
 - Unexpected entrainment of the circadian clock by the cell cycle at different FBS in NIH3T3 fibro.
- CRN model built from previous models of cell cycle and circadian clock and fitted to the data.



Kinetic Model Learning Protocols from Time Series Data

Input:

- Vector of m observed variables $\vec{x} = (x_1, \dots, x_m)$ including time x_1
- Set of n observed transitions $\{(\vec{x}_1, \vec{y}_1), \dots, (\vec{x}_n, \vec{y}_n)\}$
 1. Either **one single trace from one single initial state** $\{(\vec{x}_1, \vec{y}_1), (\vec{y}_1, \vec{y}_2), \dots, (\vec{y}_{n-1}, \vec{y}_n)\}$
 2. Or **multiple traces from multiple initial states**
 3. Possibly **branching traces**, e.g. $(\vec{x}_i, \vec{y}_j), (\vec{x}_i, \vec{y}_k)$

Output:

- Kinetic model of interactions between the observed variables (no latent variable)
 1. Either a **deterministic ODE model** $\{\dot{x}_i = f_i(\vec{x})\}_{i=1}^m$
 2. Or **CRN model** $\{R_i \xrightarrow{v_i} P_i\}_{i=1}^r$ interpreted by det. **ODE**, non-det. **CTMC**, **Petri net** or **Boolean semantics**.
 3. Particular case of influence model with forces $\{x_{j_i} \xrightarrow{v_i} x_{j_i} + x_{k_i}\}_{i=1}^p \{x_{j_i} + x_{k_i} \xrightarrow{v_i} x_{j_i}\}_{i=p+1}^m$
- Reproducing the input traces by simulation
- Plus generalizing them for different initial states
- Inferring an explainable non-overfitted **sparse model** with few reactions, i.e. f_i 's made of few terms.

3. Global Approach, e.g. SINDy Sparse Regression Algorithm

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

Reconstructs an **ODE system** from transition traces, **as weighted sums of library functions** (e.g. monomials of degree at most 2, i.e. mass action law kinetics for reactions with at most 2 reactants)

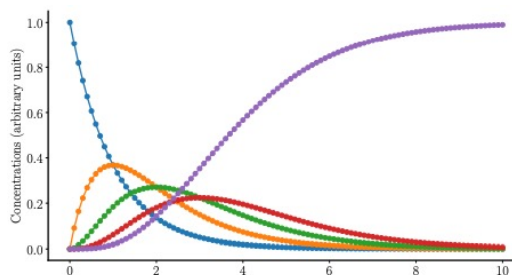
$$\mathbf{V} = \Theta(\mathbf{Y})\Xi \quad (9)$$

$\Theta(\mathbf{Y}) \in \mathbb{R}^{n \times p}$ is a library of p functions constructed from the input variables \mathbf{Y} including, for instance, first to k order polynomial interactions, e.g. $\mathbf{Y}_{\bullet,j} \odot \mathbf{Y}_{\bullet,j'}$, the sin and cos functions, e.g. $\sin(\mathbf{Y}_{\bullet,j})$, or even more sophisticated user-defined functions. The dynamics of each variable is then captured by a weighted combination of library members, the weights being encompassed in Ξ . Because it is thought that the expression of the dynamics should be sparse within the library $\Theta(\mathbf{Y})$, SINDy proposes to obtain Ξ using sparse regression.

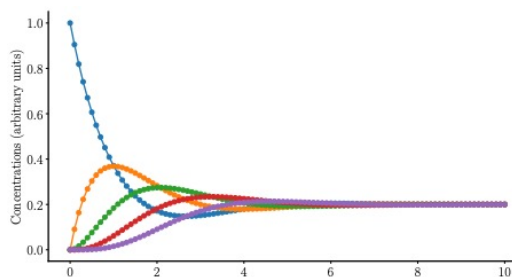
$$\Xi = \underset{\Xi \in \mathbb{R}^{p \times m}}{\operatorname{argmin}} \|\mathbf{V} - \Theta(\mathbf{Y})\Xi\|_F^2 + \lambda \|\Xi\|_1 \quad (10)$$

SINDy approximates the observed velocities and **ensures sparsity** by minimizing the sum of absolute values (Lasso regression) with **one single hyperparameter λ**

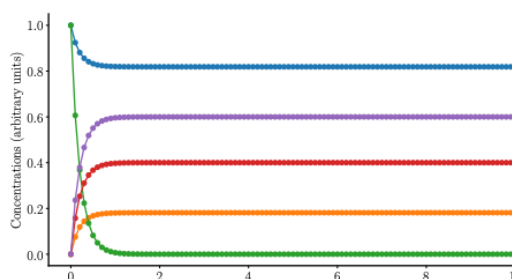
Failures of SINDy on one Single Trace from Simple Reactions



| Name | Hidden CRN | CRN inferred by Reactmine | ODE system inferred by SINDy |
|-------|-----------------------|--------------------------------------|--|
| Chain | $A \xrightarrow{1} B$ | $D \xrightarrow[1.00041]{0.99869} E$ | $\begin{cases} \dot{A} = -1.00A \\ \dot{B} = 1.00A - 1.00B \\ \dot{C} = 1.03B - 1.03C + 0.01D - 0.06AB \\ \dot{D} = 0.33B - 0.64DE \\ \dot{E} = 1.00D \end{cases}$ |
| | $B \xrightarrow{1} C$ | $C \xrightarrow[1.00071]{0.99836} D$ | |
| | $C \xrightarrow{1} D$ | $B \xrightarrow[0.99753]{1.00102} C$ | |
| | $D \xrightarrow{1} E$ | $A \xrightarrow[1.00107]{1.00069} B$ | |



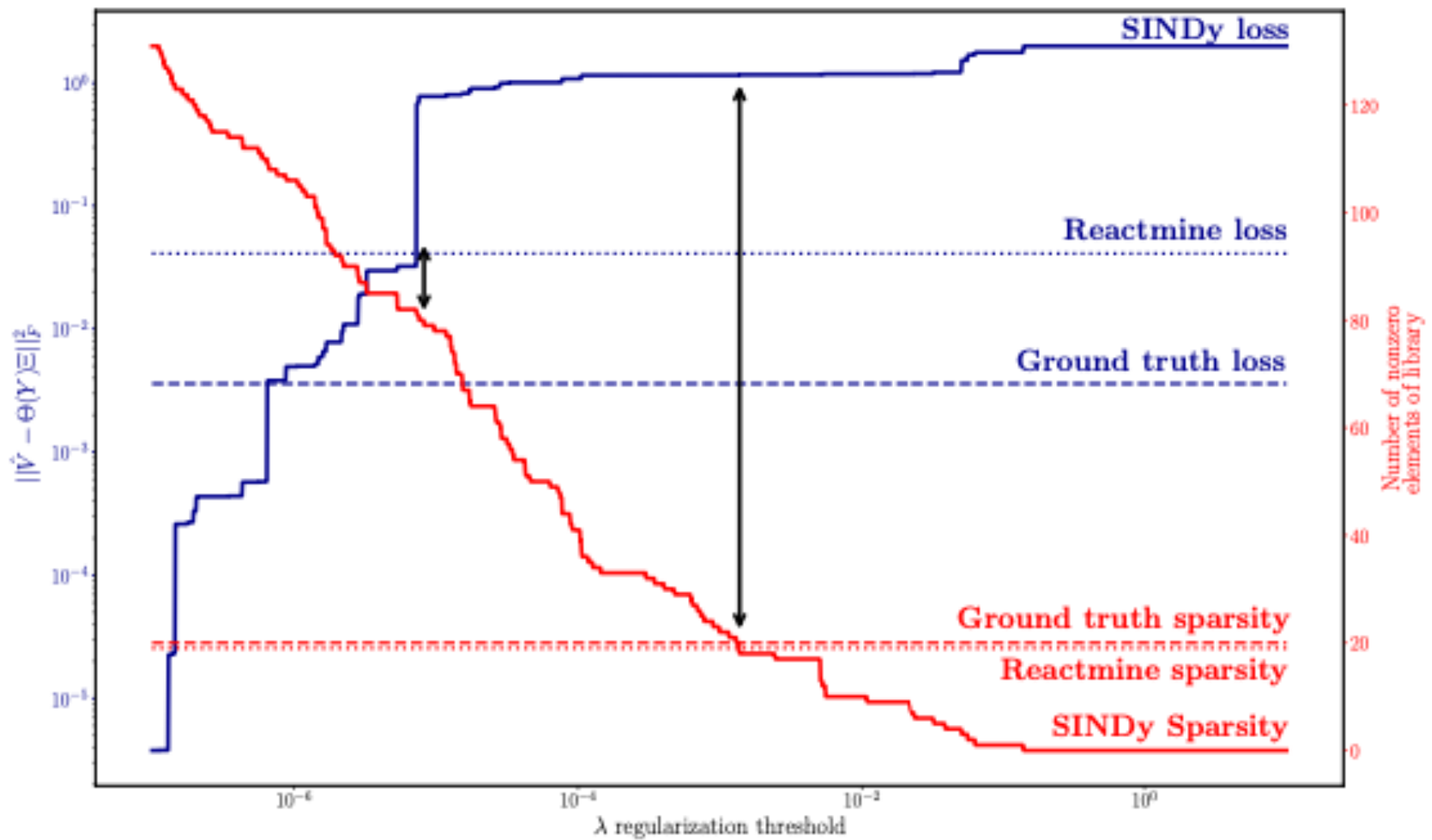
| | | | |
|------|-----------------------|--------------------------------------|---|
| Loop | $A \xrightarrow{1} B$ | $A \xrightarrow[1.00005]{0.99958} B$ | $\begin{cases} \dot{A} = -1.00A + 1.03E - 0.006D - 0.07AE - 0.06DE \\ \dot{B} = 1.00A - 1.00B + 0.004C + 0.001AB - 0.211AC - 0.092BC \\ \dot{C} = 1.14B - 1.18C - 0.002D - 0.17AB + 0.39CD \\ \dot{D} = 0.35B - 0.35E \\ \dot{E} = 0.39C + 0.457E - 4.21AE \end{cases}$ |
| | $B \xrightarrow{1} C$ | $B \xrightarrow[0.99243]{0.99748} C$ | |
| | $C \xrightarrow{1} D$ | $C \xrightarrow[0.99156]{0.99772} D$ | |
| | $D \xrightarrow{1} E$ | $D \xrightarrow[0.92045]{0.99771} E$ | |
| | $E \xrightarrow{1} A$ | $E \xrightarrow[1.00007]{0.99758} A$ | |



| | | | |
|-------------------|-------------------------------|--|---|
| Reactant Parallel | $A + C \xrightarrow{3} B + C$ | $D \xrightarrow[1.97698]{2.00109} C$ | $\begin{cases} \dot{A} = -1.12A - 510000276790.32C - 0.87AB + 1.03AD \\ \quad + 510000276788.57AC + 510000276790.32BC - 0.04DC \\ \dot{B} = 1.12A + 510000276790.24C + 0.87AB - 1.04AD \\ \quad - 510000276788.49AC - 510000276790.24BC + 0.04DC \\ \dot{C} = -2.20D + 0.02E - 0.19AD + 0.40DE \\ \dot{D} = 0.07A + 1.96D + 1.02E - 0.06DE \\ \dot{E} = -1.00E \end{cases}$ |
| | $D \xrightarrow{2} C$ | $E \xrightarrow[0.00526]{1.00101} C$ | |
| | $E \xrightarrow{1} C$ | $A + C \xrightarrow[2.95790]{2.94380} B + C$ | |

| | | | |
|------------------|-------------------------------|--|---|
| Product Parallel | $A + C \xrightarrow{1} B + C$ | $C \xrightarrow[3.12657]{2.93602} E$ | $\begin{cases} \dot{A} = -0.93C \\ \dot{B} = 0.93C \\ \dot{C} = 68364.34 - 84189.93A + 32887.73C + 14190.64D \\ \quad + 21285.97E - 17066.849AC - 40273.85AE + 990.87CD \\ \quad + 1486.24CE + 7919.89DE \\ \dot{D} = 3.73C - 1.84AC \\ \dot{E} = 65012.55A - 60751.24C - 77885.60E - 4258.48AC \\ \quad - 10214.65CE - 27068.38DE \end{cases}$ |
| | $C \xrightarrow{2} D$ | $C \xrightarrow[2.08438]{1.95735} D$ | |
| | $C \xrightarrow{3} E$ | $A + C \xrightarrow[1.04002]{0.96372} B + C$ | |

No sparse ODE system with a good loss function is inferred for any value of λ

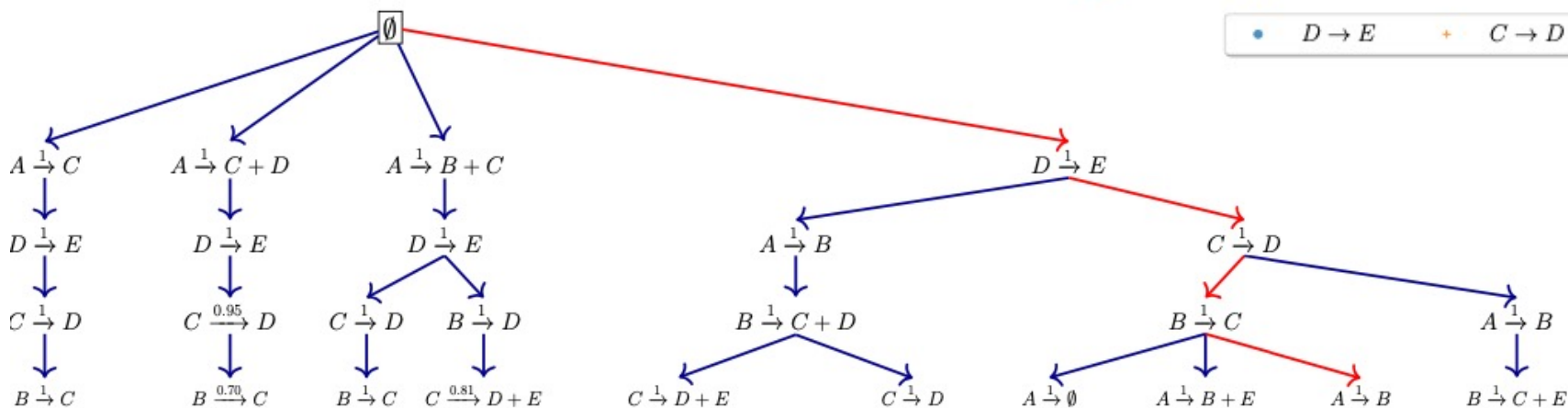
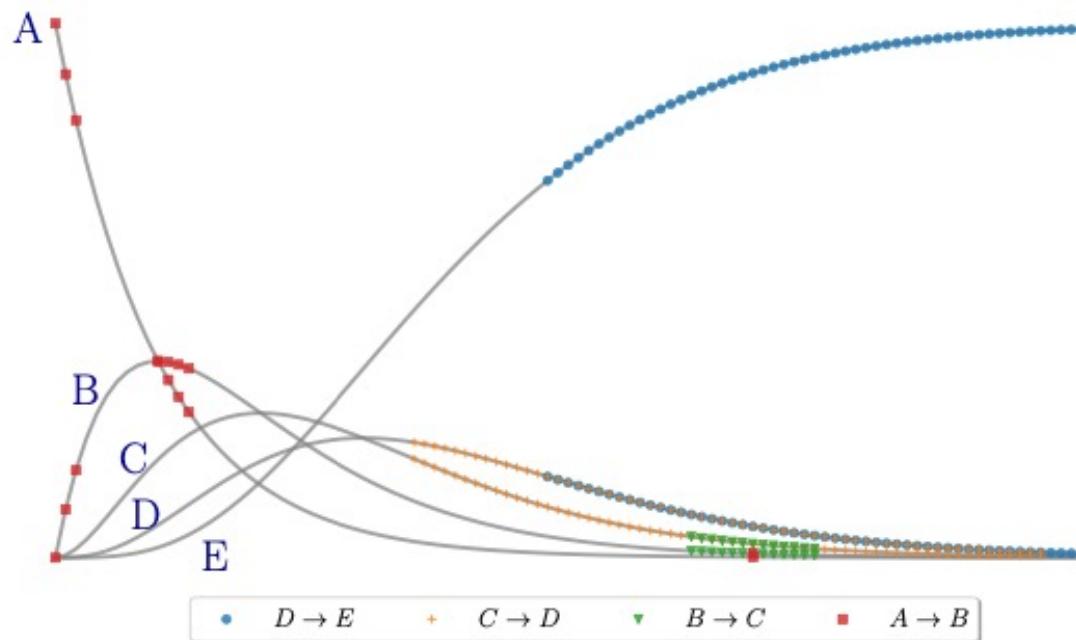


4. Local Approach: Reactmine Search Algorithm

1) Sequential inference of

- reactions with kinetics (e.g. mass action kinetics at most two reactants)
- on some transition support where the preponderant changes can be paired in one preponderant reaction.
- ranked by kinetics of lowest variance.
- trace update by subtraction of the inferred reaction

2) Final global parameter re-optimization



Reactmine: Variance-based Search Algorithm

At each node of the search tree with **bounded depth** $\leq \gamma$ (i.e. maximum number of reactions to infer)

1) For each observed transition infer a **reaction skeleton** on **concentration changes with highest velocities**

$$i_l^{\max} = \underset{i}{\operatorname{argmax}} |\tilde{v}_{l,i}| \quad \tilde{v}_{l,i} = \frac{\hat{v}_{l,i}}{\max_{1 \leq k \leq n} |\hat{v}_{k,i}|} \quad R_\delta(t_l) = \left\{ i \in \{1, \dots, m\}, \hat{v}_{l,i} < 0, \left| \frac{\hat{v}_{l,i}^{\max}}{\hat{v}_{l,i}} \right| \leq \delta \right\} \text{ with velocity ratio } \leq \delta \leq \delta_{\max}$$

$$P_\delta(t_l) = \left\{ i \in \{1, \dots, m\}, \hat{v}_{l,i} > 0, \left| \frac{\hat{v}_{l,i}^{\max}}{\hat{v}_{l,i}} \right| \leq \delta \right\}$$

2) Rank reaction candidates (k, R, P) according to **kinetic variance** $\leq \alpha$ $\hat{\sigma}_j = \frac{1}{|T(r)|} \sum_{l \in T(r)} \left(\frac{|v_{l,j}|}{\prod_{i \in R} y_{l,u}} - \hat{k}_j \right)^2$
on their supporting transitions $T(r) = \{l \in \{1, \dots, n\} \mid \exists \delta \in [1, \delta_{\max}], r^\delta(t_l) = (R, P)\}$

3) Create successor nodes for $\leq \beta$ **best kinetic variance candidates** and **subtract reaction effect on the trace**

4) Finally, optimize the inferred kinetic parameters on the whole trace $\mathbf{k} = \underset{\mathbf{k} \in \mathbb{R}_+^p}{\operatorname{argmin}} \|\mathbf{V} - \mathbf{F}(\mathbf{Y}, \mathbf{k})\mathbf{S}\|_F^2$

Four parameters: $\gamma, \delta_{\max}, \alpha, \beta$.

Computational Complexity of Reactmine

4 hyperparameters $\alpha, \beta, \gamma, \delta_{max}$

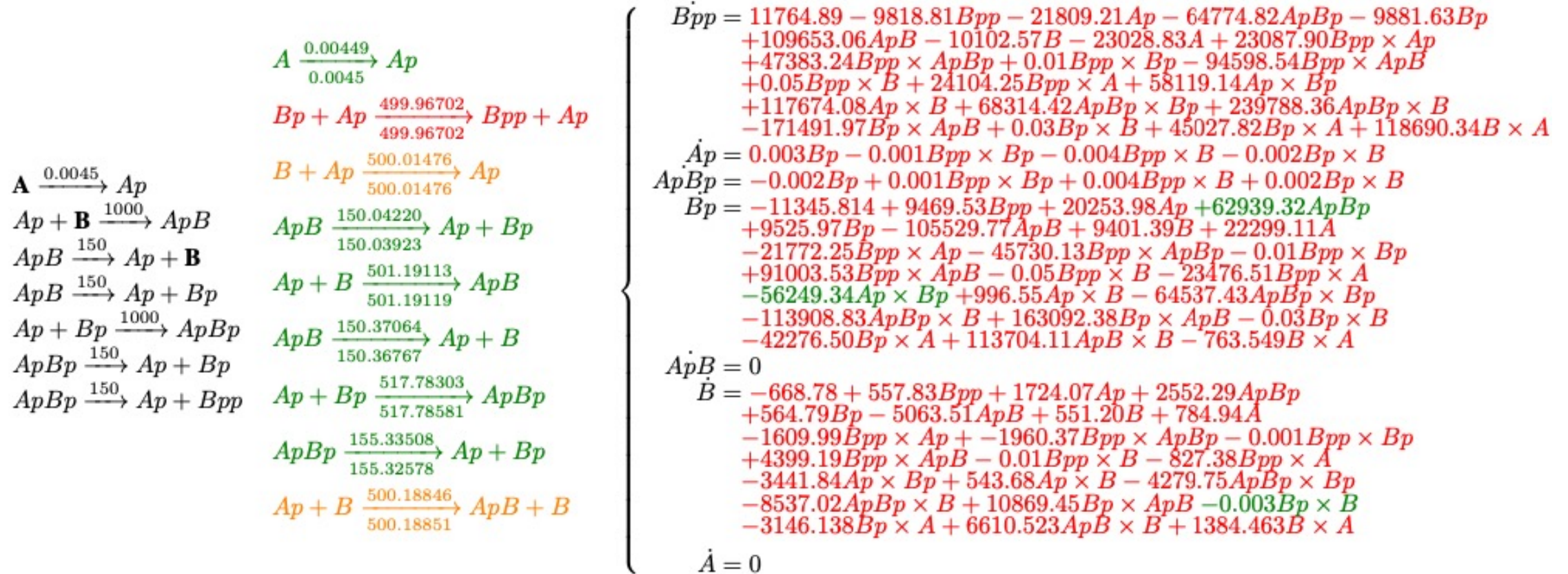
n species, m observed transitions.

Proposition. The computational time complexity to infer one reaction (R, P, f) is $\mathcal{O}(nmJ)$ where n is the number of time points, m the number of species, and $J = |R \cup P|$. The global time complexity of Reactmine is $\mathcal{O}(\beta^\gamma nmJ)$.

Proof. Inferring the reaction kinetics constant involves the computation of a mean for each species present in the reaction (Equation 4), which is $\mathcal{O}(nJ)$. In the worst-case, a lookup for a catalyst species is necessary, at a cost of $\mathcal{O}(nJm)$. The update of velocities performed in Equation 6 is $\mathcal{O}(nJ)$. Generating reaction skeletons requires the computation of the species displaying highest absolute variations for each time point, which is $\mathcal{O}(nm)$. After that, the sets $R_\delta(i)$ and $P_\delta(i)$ are obtained with a bounded number of δ values. The time complexity for the inference of one reaction is therefore $\mathcal{O}(nmJ)$.

Since the depth of the search tree is bounded by γ and each node has at most β children, the time complexity of Reactmine is thus $\mathcal{O}(\beta^\gamma nmJ)$.

Results of Reactmine and SINDy on MAPK cell signaling CRN



- Reactmine succeeds in reproducing traces with algebraically equivalent reactions and 1 approximation
- SINDy fails not only on sparsity but also on reproducing simulation traces.

Results of Reactmine on Videomicroscopy Data

Discovery of $G2 \rightarrow \text{Rev}$ and $G1 \rightarrow G2$ as influences with max mean effect

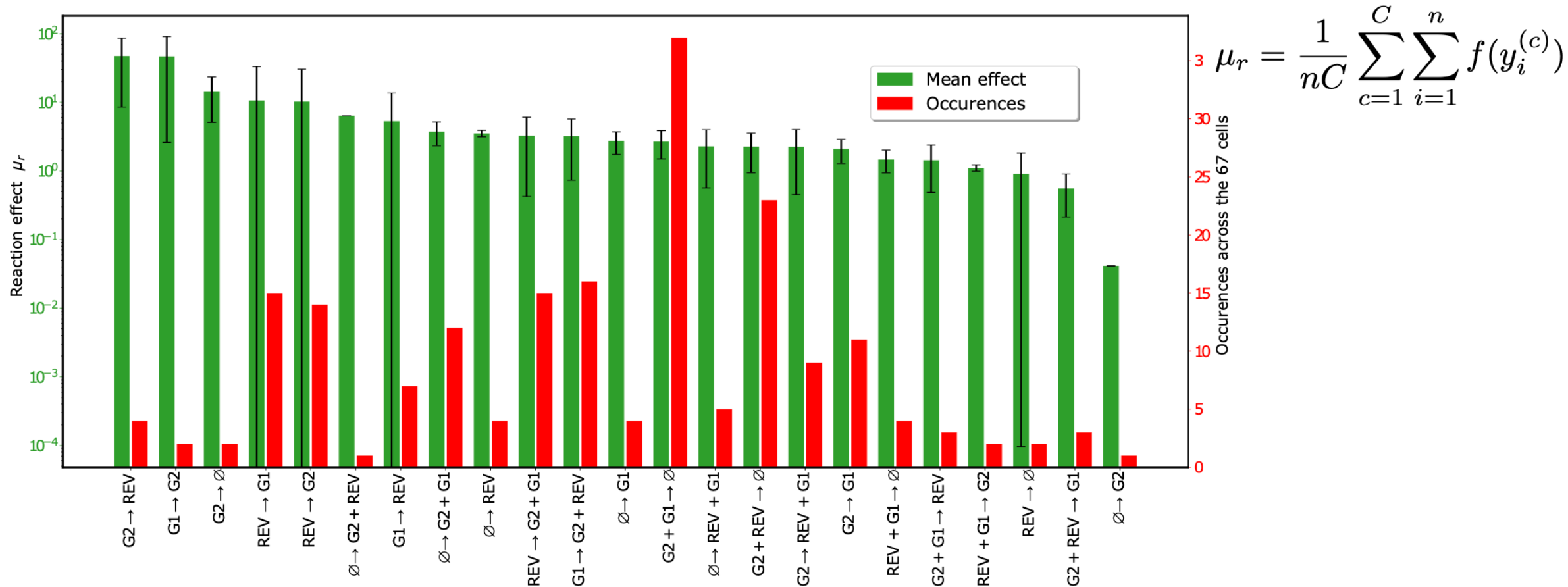


Figure 3: **Results of Reactmine on fluorescence videomicroscopy dataset.** **a** Plot of 3 cells among the 67 cells of the videomicroscopy data. Traces have been smoothed with a moving average. **b** Ranking of mean effect (green) and occurrences (red) of inferred reactions on 67 cells. Bars in green report the mean effect of each reaction across time and cells in the videomicroscopy data. Standard deviation is computed across the cells.

5. Correctness of Lasso Regression [Yu Zhao 2006]

For each sample, The Lasso estimates $\hat{\beta}^n(\lambda)$ are defined by $\hat{\beta}^n(\lambda) = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \{ \|Y_n - X_n \beta\|_2^2 + \lambda \|\beta\|_1 \}$

For $\lambda = 0$, if X_n of full rank we have a closed form solution $\hat{\beta} = (X_n^T X_n)^{-1} X_n^T Y_n$

Let $X_n(1), X_n(2)$ be the columns of X_n multiplied by non-zero/zero β coefficients and $C^n = \frac{1}{n} X_n X_n^T$

relevant/irrelevant

$$C^n = \begin{pmatrix} C_{11}^n & C_{12}^n \\ C_{21}^n & C_{22}^n \end{pmatrix}$$

The Strong irrepresentable Condition $|C_{21}^n (C_{11}^n)^{-1} \operatorname{sign}(\beta_{(1)}^n)| < 1$

expresses low correlations between the predictors in the model and not in the model

Theorem [Yu Zhao 2006] Sufficient condition for **correctness of Lasso model selection with exp rate in n**

For fixed q and p , and under regularity conditions (1) and (2), Lasso is strongly sign consistent if the Strong Irrepresentable Condition holds. That is, when Strong Irrepresentable Condition holds, $\forall \lambda_n$ that satisfies $\frac{\lambda_n}{n} \xrightarrow{n} 0$ and $\frac{\lambda_n}{n^{(1+c)/2}} \xrightarrow{n} +\infty$ with $0 < c < 1$, we have :

$$\mathbb{P}(\hat{\beta}^n(\lambda_n) =_s \beta) = 1 - \mathcal{O}(\exp(-n^c))$$

Where the conditions (1) and (2) are : $C^n \xrightarrow{n} C$ $\frac{1}{n} \max_{1 \leq i \leq n} ((x_i^n)^\top x_i^n) \xrightarrow{n} 0$

Multi-traces on a Combinatorics of Initial Conditions

To make SinDy work, we concatenate m traces stemming from different initial conditions. The resulting input matrix when the library function is the products of concentrations is of the form:

$$\Theta(X_n) = \begin{bmatrix} x_1^1(0) & x_2^1(0) & \cdots & x_1^1(0)x_2^1(0) & \cdots \\ x_1^1(\Delta t) & x_2^1(\Delta t) & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots & \cdots \\ x_1^1((n-1)\Delta t) & x_2^1((n-1)\Delta t) & \cdots & \cdots & \cdots \\ x_1^2(0) & x_2^2(0) & \cdots & x_1^2(0)x_2^2(0) & \cdots \\ \vdots & \vdots & \ddots & \cdots & \cdots \\ x_1^m((n-1)\Delta t) & x_2^m((n-1)\Delta t) & \cdots & \cdots & \cdots \end{bmatrix}$$

where: $x_k^i(l\Delta t)$ is the concentration of species k at time $l\Delta t$ on the i -th trace.

n is the number of time points

n_s is the number of species (used later)

Δt is the time step

Graphical Conditions for Correctness of Lasso Regression

Type 1 CRN: **Atmost One Reactant per Reaction on Multitraces with 1 Non-zero**

Theorem

For a CRN of type 1, suppose we concatenate the traces as above and choose the optimal library function. Suppose also that the time step is such that $\Delta T =_{n \rightarrow \infty} o(\frac{1}{n})$. Then Zhao's irrepresentable condition holds for a big enough n where n is the number of time points.

Proof

$$\begin{aligned}
 C_{1,1}^n(i, j) &= \frac{1}{nm} \sum_{k=1}^m \sum_{l=0}^{n-1} X_i^k(l\Delta T) X_j^k(l\Delta T) & \frac{1}{n} \langle X_i^k, X_j^k \rangle &= \frac{1}{n} \sum_{l=0}^{n-1} X_i^k\left(\frac{l}{n}\right), X_j^k\left(\frac{l}{n}\right) \text{ by Cesaro lemma} \\
 &= \frac{1}{m} \sum_{k=1}^m \frac{1}{n} \langle X_i^k, X_j^k \rangle & &\xrightarrow{n \rightarrow \infty} X_i^k(0), X_j^k(0) \\
 &= \mathbb{E}_m \left[\frac{1}{n} \langle X_i^k, X_j^k \rangle \right] & C_{1,1}^n(i, j) &\xrightarrow{n \rightarrow \infty} \mathbb{E}_m [X_i^k(0) X_j^k(0)] \text{ by dominated convergence theorem} \\
 C_{1,1}^n &\xrightarrow{n \rightarrow \infty} I_p & (C_{1,1}^n)^{-1} &\xrightarrow{n \rightarrow \infty} I_p \\
 C_{2,1}^n(i, j) &\xrightarrow{n \rightarrow \infty} \mathbb{E}_m [X_{a_i}^k(0) X_{b_j}^k(0)] & C_{2,1}^n(i, j) &\xrightarrow{n \rightarrow \infty} 0 \quad |C_{2,1}^n (C_{1,1}^n)^{-1} \beta^1| < 1 \text{ for large enough } n
 \end{aligned}$$

Type 2 CRN: At most 2 Reactants per Reaction on Multitraces with 2 Non-zero Initial Conditions

Multitraces with at most two non-zero initial conditions

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}$$

if $C^n(i, j)$ has 3 or 4 species :

$$C_{i,j}^n \xrightarrow{n \rightarrow \infty} \mathbb{E}_m \left[X_{a_i}^k(0) X_{b_i}^k(0) X_{a_j}^k(0) \right] = 0$$

if $C^n(i, j)$ has 2 species :

$$C_{i,j}^n \xrightarrow{n \rightarrow \infty} \mathbb{E}_m \left[X_{a_i}^k(0)^{\epsilon_1} X_{a_j}^k(0)^{\epsilon_2} \right] = \frac{1}{m} = \frac{2}{n_s(n_s + 1)} \quad \epsilon_1, \epsilon_2 \in (1, 2)$$

if $C^n(i, j)$ has 1 specie :

$$C_{1,1}^n = \begin{bmatrix} A & D^\top \\ D & \frac{1}{m} I_{\frac{n_s(n_s-1)}{2}} \end{bmatrix}$$

Non-sparse covariance matrix

Yu & Zhao's correctness condition may be not satisfied

SINDy on Multiple Traces with 2 Non-zero Initial Conditions

| Ground Truth | | Simple trace | | Multiple trajectories | |
|-------------------|------|---------------------|------|-----------------------|------|
| Equation | Rate | Equation | Rate | Equation | Rate |
| $A \rightarrow B$ | 1 | $A \rightarrow B$ | 1.00 | $A \rightarrow B$ | 1.00 |
| $B \rightarrow C$ | 1 | $B \rightarrow C$ | 1 | $B \rightarrow C$ | 1.00 |
| $C \rightarrow D$ | 1 | $C \rightarrow D$ | 0.48 | $C \rightarrow D$ | 1.00 |
| $D \rightarrow E$ | 1 | $D + E \rightarrow$ | 1.02 | $D \rightarrow E$ | 1.06 |

| Ground Truth | | Simple trace | | Multiple trajectories | |
|-------------------|------|-------------------|------|-----------------------|------|
| Equation | Rate | Equation | Rate | Equation | Rate |
| $A \rightarrow B$ | 1 | $A \rightarrow B$ | 1.00 | $A \rightarrow B$ | 1.00 |
| $B \rightarrow C$ | 1 | $B \rightarrow C$ | 1.00 | $B \rightarrow C$ | 1.00 |
| $C \rightarrow D$ | 1 | $C \rightarrow D$ | 1.00 | $C \rightarrow D$ | 1.00 |
| $D \rightarrow E$ | 1 | $D \rightarrow E$ | 1.00 | $D \rightarrow E$ | 1.00 |
| $E \rightarrow A$ | 1 | $E \rightarrow A$ | 1.00 | $E \rightarrow A$ | 1.00 |

| Ground Truth | | Simple trace | | Multiple trajectories | |
|-------------------|------|-------------------|------|-----------------------|------|
| Equation | Rate | Equation | Rate | Equation | Rate |
| $A \rightarrow B$ | 1 | $A \rightarrow B$ | 1.01 | $A \rightarrow B$ | 1.01 |
| $D \rightarrow C$ | 2 | | | $D \rightarrow C$ | 1.99 |
| $E \rightarrow C$ | 3 | | | $E \rightarrow C$ | 3.01 |

| Ground Truth | | Simple trace | | Multiple trajectories | |
|-------------------|------|-----------------------|------|-----------------------|------|
| Equation | Rate | Equation | Rate | Equation | Rate |
| $A \rightarrow B$ | 1 | | | $A \rightarrow B$ | 1.01 |
| $C \rightarrow D$ | 2 | | | $C \rightarrow D$ | 2.04 |
| $C \rightarrow E$ | 3 | $C + D \rightarrow E$ | 0.75 | $C \rightarrow E$ | 3.06 |

| Ground Truth | | Single trace | | Multiple trajectories | |
|-----------------------|------|------------------------|------|-----------------------|------|
| Equation | Rate | Equation | Rate | Equation | Rate |
| $B + D \rightarrow E$ | 0.3 | | | $B + D \rightarrow E$ | 0.3 |
| $2A \rightarrow B$ | 0.1 | $2A \rightarrow B + C$ | 0.20 | $2A \rightarrow B$ | 0.1 |
| $A \rightarrow C$ | 0.2 | $A + E \rightarrow$ | 0.16 | $A \rightarrow C$ | 0.2 |
| $C \rightarrow D$ | 0.13 | $C + D \rightarrow A$ | 0.09 | $C \rightarrow D$ | 0.13 |

Conclusion

Global approaches such as SINDy, using sparse regressions such as LASSO, are correct if the library terms are uncorrelated [You Zhao 2006]

- **Theorem:** conditions satisfied **for at most monoreactant systems** (e.g. positive influence systems), on data about the **first transitions** observed from **many initial states with one non 0 variable**
- First transitions observed from many initial states with few non-zero values are more informative already seen for PAC learning of Boolean influence models [Carcano Fages Soliman CMSB 2017]
- Multitraces from sparse initial states improve results but are unrealistic in biology

Local approaches such as **Reactmine** **search for preponderant reactions in some supporting transitions**

- Single trace multireactant experiments: successes on synthetic time series over a few variables
- Absence of inference of latent variables is a strong limitation
- Limited to deterministic ODE-CRN dynamics

Future Work on Reactmine

- Periodic reoptimization of the kinetic parameters of inferred reactions rather than final reoptimization ?
- Use beam search (i.e. A* with bounded frontier) to scale up beyond 7 reactions?
- Infer latent variables on contradictory (non-deterministic) transitions?
 - Videomicroscopy data are limited to a few markers
 - Latent variables are thus needed in the CRN model to reproduce the traces
- Consider stochastic interpretation of CRN (Continuous Time Markov Chain semantics)?
 - To infer CRN models of branching processes
 - E.g. infer model of cell differentiation from RNAseq data

Reactmine Computation Times with Hyperparameter Gridsearch

| CRN | Hyperparameters (α , β , δ_{\max} , γ) | Computation time in seconds | Number of inferred CRNs | Grid search values for (α , β , δ_{\max} , γ) | Computation time with grid search |
|-------------------|--|--------------------------------|----------------------------|---|--------------------------------------|
| Chain | (0.02, 7, 3, 4) | 0.31 | 128 | $\alpha \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.5\}$ $\beta \in \{4, 5, 6, 7, 8\}$ $\delta_{\max} = 3$ $\gamma \in \{3, 4, 5, 6\}$ | 3035.95 |
| Loop | (0.02, 7, 3, 6) | 40.78 | 4198 | $\alpha \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.5\}$ $\beta \in \{4, 5, 6, 7, 8\}$ $\delta_{\max} = 3$ $\gamma \in \{3, 4, 5, 6\}$ | 30643.19 |
| Reactant-Parallel | (0.02, 4, 3, 5) | 1.81 | 265 | $\alpha \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.5\}$ $\beta \in \{4, 5, 6, 7, 8\}$ $\delta_{\max} = 3$ $\gamma \in \{3, 4, 5, 6\}$ | 970.23 |
| Product-Parallel | (0.02, 5, 3, 3) | 0.13 | 24 | $\alpha \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.5\}$ $\beta \in \{4, 5, 6, 7, 8\}$ $\delta_{\max} = 3$ $\gamma \in \{3, 4, 5, 6\}$ | 334.06 |
| MAPK | (0.02, 7, 3, 10) | 3012.85 | 33104 | $\alpha \in \{0.0025, 0.005, 0.0075, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.1, 0.2\}$ $\beta \in \{6, 8, 10\}$ $\delta_{\max} = 3$ $\gamma \in \{7, 8, 9, 10\}$ | 77179.92 |