Sparse Estimation in Ordinary Differential Equation Systems

Frederik Riis Mikkelsen
Department of Mathematical Sciences
University of Copenhagen
Copenhagen 2100, Denmark
Email: frm@math.ku.dk

Niels Richard Hansen
Department of Mathematical Sciences
University of Copenhagen
Copenhagen 2100, Denmark
Email: niels.r.hansen@math.ku.dk

***Abstract*—Understanding chemical reaction networks in systems biology is an important dicipline, as it provide the means for quantifying downstream effects of chemical or medical interventions. Thus identifying these systems from noisy data is a major challenge with far reaching applications – such as to the inference of phosphoprotein interaction networks as in [4].**

For mass action kinetics the system structure – the network – is encoded via sparsity in a parameter vector, whose dimension increases rapidly with the number of species. We have developed an algorithm for simultaneous system identification and parameter estimation via minimisation of a penalised loss function. The global minimiser is difficult to find, and focus has been on computational aspects, as well as variance-reduction techniques.

Several techniques are combined to cope with the computational and statistical aspects. The resulting method is implemented in an R-package, which provides sparse estimates of systems with up to 10^5 reactions.

I. Introduction

Consider d chemical components (i.e., NaCl, H_2O, phosphoproteins, etc.), denoted $x = (x_l)_{l=1}^d$, whose abundances are governed by p chemical reactions:

$$a_{j1}x_1 + ... + a_{jd}x_d \xrightarrow{k_j} b_{j1}x_1 + ... + b_{jd}x_d, \quad j = 1, ..., p, \quad (1)$$

where $A = (a_{j,l})_{j,l}$ and $B = (b_{j,l})_{j,l}$ are non-negative stoichiometric coefficients and $k = (k_j)_j$ are non-negative reaction rates. According to the law of mass action, the vector x of abundances, can be approximated by solving the ordinary differential equation (ODE):

$$x(t, k) = x_0 + \int_0^t (B - A)^{\mathsf{T}} \mathrm{diag}(x(s, k)^A) k \; ds, \quad t \in R, \quad (2)$$

where $x_0 \in R^d$ is the initial abundance and $x^A = (\prod_{l=1}^d x_l^{a_{j,l}})_j$.

We observe the x-process at n time points, $y(t_i) = x(t_i, k^{s(i)}) + \xi_i$, with $(\xi_i)_{i=1}^n$ i.i.d. noise and $s(i) \in \{1, ..., S\}$ denoting experimental condition (e.g., knockdowns, stimuli, etc.), where $k^{s(i)} = k \circ c_{s(i)}$ (Hadamard product of baseline rate and non-negative scales). We infer the reaction network from a large list of possible reactions by enforcing sparsity in the estimation of k. The penalised squared error loss function is considered:

$$\ell(x_0, k) = \frac{1}{2} \sum_{i=1}^n \sum_{l=1}^d w_{i,l}(y_l(t_i) - x_l(t_i, k^{s(i)}))^2 + \lambda \sum_{j=1}^p v_j \mathrm{pen}(k_j) \quad (3)$$

with pen a sparsity enforcing penalty, e.g., elastic net, SCAD, MCP, and $w = (w_{i,l})_{i,l}$, $v = (v_j)_j$ are observational and penalty weights.

II. Computational aspects

Minimising (3) is a high dimensional problem, as the number of possible reactions, p, grows quickly in the number of chemical species, d. For example there are $p = d^2(d-1)$ distinct simple enzymatic reactions of the form

$$x_{l_1} + x_{l_2} \to x_{l_1} + x_{l_3}. \quad (4)$$

Moreover, evaluating the derivative of (3) requires solving another ODE system taking values in $R^{d(d+p)}$, called the sensitivity equations (see [7]). Though numerical solvers of ODEs are fundamentally sequential in nature the sensitivity equations can be solved in parallel. Yet these evaluations should be kept at a minimum.

We propose combining a proximal-gradient based method (see [1], [8]) with occasional screening for strong coordinates to reduce the number of full gradient evaluations. Furthermore, good initialisations of k are provided by integral matching (see [6] for details). Figure 3 shows the solution recovered from a small simulated mass action kinetics system.

III. Variance reduction and re-weighting

The loss function (3) is likely to have a vast number of local minima, primarily due to nonlinearity of x. Consequently, extra variance of the minimal loss estimator is introduced. In order to reduce this, penalty regularization will not alone suffice.

Additionally, if the reactions operate on different scales, appropriate adjustments of $(v_j)_j$ and $(c_{s(i)})_i$ are required. For linear least squares problems, this is usually handled by standardising the columns of the design, but no immediate equivalent exists for minimising (3). However, the preliminary integral matching procedure replaces (3) with a (squared) surrogate loss function using non-parametric estimates of the x-process (similar to methods of [2], [3]). Hence classic tools for standardising and reweighting ([9]) the system through w, v and $(c_{s(i)})_i$ are applicable.

Moreover, whole families of non-parametric estimators of x are applied to give multiple initialisations to minimising (3), which through the continuation principle produces solution paths over the tuning parameter λ. Collectively the solutions paths reduce the issue of multiple minima and, through model averaging procedures and/or stability selection, also reduce the variance of the final estimator.

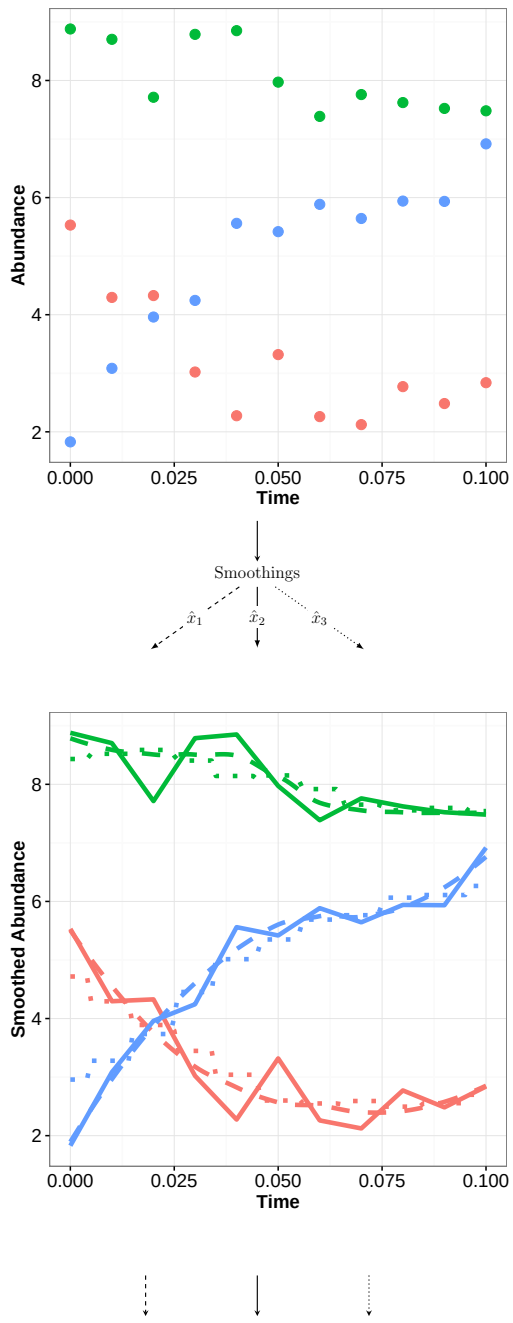
IV. Results

The method summarised above is implemented in an R-package, which can handle high-dimensional reaction systems (e.g., systems on the form (1) with up to $p = 10^5$) using either elastic net, SCAD and MCP penalties. The package centers around two main features: MAKER (Mass Action Kinetics Estimation with Regularization) and AIM (Adaptive Integral Matching). AIM contains the tools for automatically producing initial estimates, adapting weights and scales. These are then pipelined to MAKER, which runs the proximal gradient algorithm with screenings. See Figures 1 and 2 for overviews of the algorithms.

References

[1] Bolte et al. *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program. (2014) 146(1), 146–459.

[2] Brewer et al., *Fitting ordinary differential equations to short time course data*, Phil. Trans. R. Soc. A (2008) 366, 519–544.



AIM:

$$X_i^{\hat{x}} \coloneqq \int_{t_0}^{t_i} (B - A)^T \mathrm{diag}(\hat{x}(t)^A \circ c_{s(i)})\, dt, \quad i = 1, ..., n$$

$$\hat{k}^{\hat{x}} \coloneqq \arg\min_k \tfrac{1}{2} \sum_{i=1}^{n} \|y(t_i) - \hat{x}(t_0) - X_i^{\hat{x}} k\|_2^2 + \lambda \sum_{j=1}^{p} v_j \mathrm{pen}(k_j)$$

$$c_{s(i)} \propto \left(\frac{1}{\sum_{\hat{x}} \|(X_{i,j}^{\hat{x}})_i\|_2} \right)_j \circ c_{s(i)}, \quad v \propto \left(\frac{1}{\sum_{\hat{x}} |\hat{k}_j^{\hat{x}}|} \right)_j \circ v$$

Fig. 1. AIM: Preliminary processesing of data to produce automatic initial parameter values and adapting scales and weights to those.

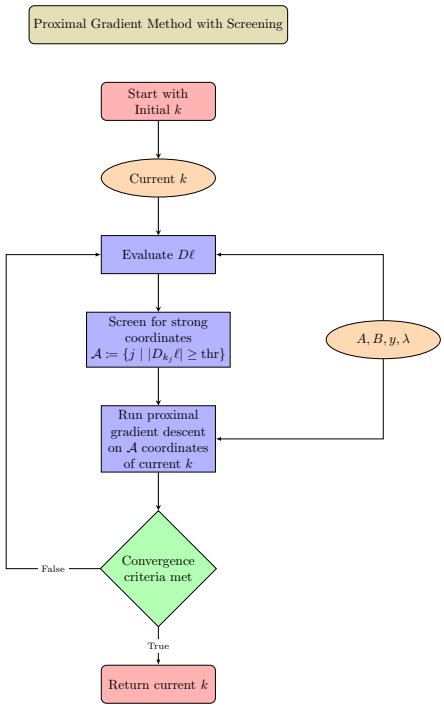


Fig. 2. Flowchart of proximal gradient method mixed with occasional screening used in MAKER.

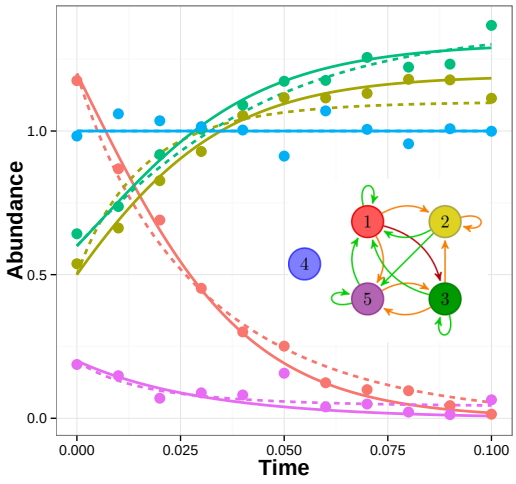


Fig. 3. Simulated mass action kinetics of enzymatic reactions of the form (4) with $d = 5$ and $p = 100$. Inferred network (green = TP, red = FP, orange = FN). Fully drawn lines are true ODE solutions, dashed lines are least squares estimates using inferred network.

[3] N.J-B Brunel, *Parameter estimation of ODEs via nonparametric estimators*, Electronic Journal of Statistics, (2008) 2, 1242–1267.

[4] Hill et al., *Inferring causal molecular networks: empirical assessment through a community-based effort*, Nat. Meth. (2016) 13(4), 310–318.

[5] J. Kennedy and R. Eberhart, *Particle Swarm Optimization*, Proceedings of IEEE International Conference on Neural Networks. IV., 1942–1948.

[6] F. R. Mikkelsen, *Computational Aspects of Parameter Estimation in Ordinary Differential Equation Systems*, Proceedings of the 19th European Young Statisticians Meeting (2015), 94–99.

[7] Ramsay et al., *Parameter estimation for differential equations: a generalized smoothing approach*, J R Stat Soc Series B Stat Methodol (2007), 69(5), 741–796.

[8] S. Yun and K-C Toh, *A coordinate gradient descent method for l1-regularized convex minimization*, Computational Optimization and Applications (2011), 48(2), 273–30.

[9] H. Zou, *The Adaptive Lasso and Its Oracle Properties*, Journal of the American Statistical Association, 101:476, 1418–1429.