



## Sparse Estimation in Ordinary Differential Equation Systems

Mikkelsen, Frederik Vissing; Hansen, Niels Richard

*Publication date:*  
2017

*Document version*  
Peer reviewed version

*Citation for published version (APA):*  
Mikkelsen, F. V., & Hansen, N. R. (2017). *Sparse Estimation in Ordinary Differential Equation Systems*. Paper presented at Signal Processing with Adaptive Sparse Structured Representations, Lisbon, Portugal.

# 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**—Ordinary differential equation systems are used to model abundances of chemical species as deterministic continuous time processes. Identification of the system from noisy data is a major challenge with applications in systems biology – such as to the inference of phosphoprotein interaction networks as in [3].

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 handles enzymatic systems of up to 100 chemical species.

## I. INTRODUCTION

We consider a deterministic continuous time process  $x$  taking values in  $R^d$  governed by an ordinary differential equation (ODE) system. More precisely, we assume that  $x$  solves the initial value problem:

$$x(t, \theta) = x_0 + \int_0^t f(x(s, \theta), \theta) ds, \quad t \in R, \quad x_0 \in R^d \quad (1)$$

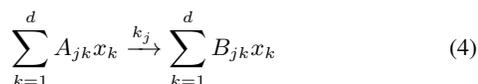
where  $f : R^d \times R^p \rightarrow R^d$  determines the ODE system and  $\theta \in R^p$  denotes the parameter of interest. Let  $y = (y_i)_{i=1}^n$  be discrete observations from the process at distinct time points  $y_i = x(t_i) + \xi_i$ , with  $(\xi_i)_{i=1}^n$  i.i.d. noise. The  $\ell^1$ -penalised squared error loss function then reads:

$$\ell(\theta) = \frac{1}{2} \sum_{i=1}^n \left\| y_i - x_0 - \int_0^{t_i} f(x(s, \theta), \theta) ds \right\|_2^2 + \lambda \sum_{j=1}^p w_j |\theta_j| \quad (2)$$

Our focus is on the class of ODE models called mass action kinetics models, where  $f$  is on the form

$$f(x, \theta) = (B - A)^T \text{diag}(x^A) \theta, \quad x \in R_+^d, \quad \theta \in R_+^p. \quad (3)$$

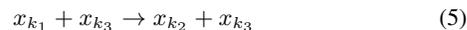
Each row of the so-called stoichiometric matrices  $A$  and  $B$  corresponds to a reaction:



with rate  $k_j \geq 0$ , and the reaction only contributes to the system if  $k_j > 0$ . If the stoichiometric matrices contain an exhaustive list of reactions that would mimic most common chemical behaviours, obtaining a sparse estimate of  $\theta$  through minimising (2) would therefore present a valid solution to the network identification problem.

## II. COMPUTATIONAL ASPECTS

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



Moreover, evaluating the differential of (2) requires solving another ODE system taking values in  $R^{d(d+p)}$ , called the sensitivity equations (see [5]). 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 [6]) with occasional screening for strong coordinates to reduce the number of full gradient evaluations. Furthermore, the tools proposed in [1] and [2] combined with the linearity of  $\theta \mapsto f(x, \theta)$  in (3) are used to provide a good initialisation of  $\theta$  for the minimisation. See Figures 1 and 4. Figure 3 shows the solution recovered from a small simulated mass action kinetics system.

## III. VARIANCE REDUCTION AND RE-WEIGHTING

The loss function (2) is likely to have a vast number of local minima, primarily due to nonlinearity of  $f$ . Consequently, extra variance of the minimal loss estimator is introduced, and the  $\ell^1$  regulation will not alone suffice in reducing it.

Additionally, in order for the  $\ell^1$ -penalty to regularise each reaction equally, the weights  $(w_j)_{j=1}^p$  must reflect both the absolute magnitude and the order at which the parameters enters the model. For linear least squares problems, this is usually handled by standardising the columns of the design, but no immediate equivalent exists for minimising (2).

We propose combining the ideas of multiple particle search algorithms, [4], with adaptive re-weighting schemes, [7], in order to cope with both issues. Moreover, these two schemes are incorporated into the continuation principle for minimising (2) for a sequence of different tuning parameters, see Figure 2.

## IV. RESULTS

The method summarised above is implemented in an R-package, which handles systems of (5) with up to  $d = 100$ . Furthermore, the package offers other regularisations, such as  $\ell^2$ , SCAD and MCP, and also handles missing coordinates of the observations.

## REFERENCES

- [1] Brewer et al., *Fitting ordinary differential equations to short time course data*, Phil. Trans. R. Soc. A (2008) 366, 519544.
- [2] N.J-B Brunel, *Parameter estimation of ODEs via nonparametric estimators*, Electronic Journal of Statistics, (2008) 2, 12421267.
- [3] Hill et al., *Inferring causal molecular networks: empirical assessment through a community-based effort*, Nat. Meth. (2016) 13(4), 310-318.

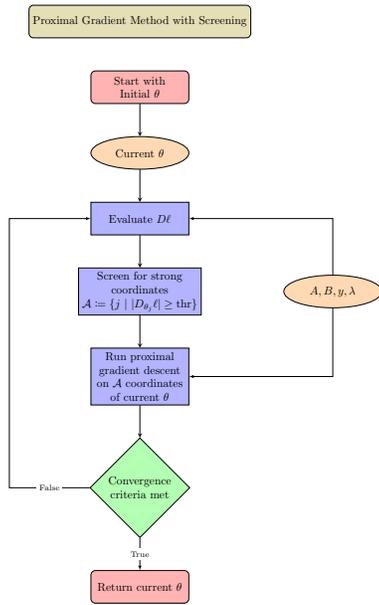


Fig. 1. Flowchart of proximal gradient method mixed with occasional screening

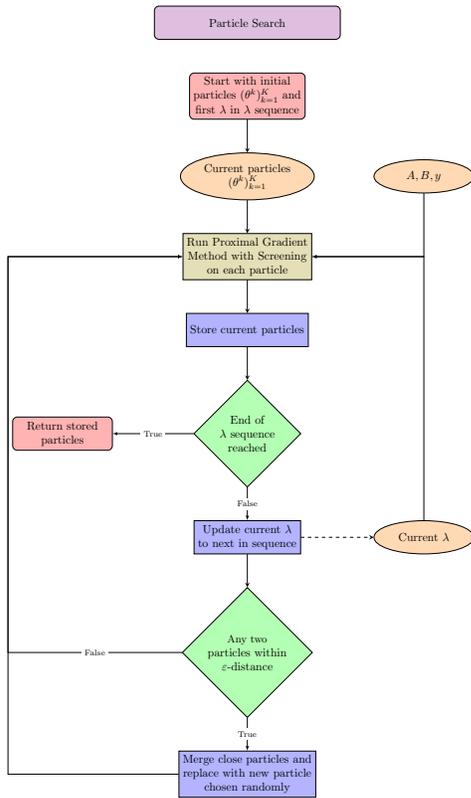


Fig. 2. Flowchart of particle method.

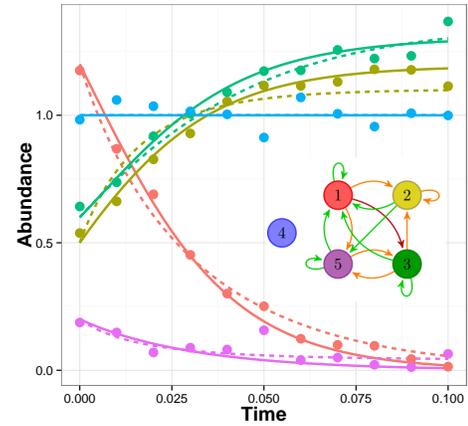


Fig. 3. Simulated mass action kinetics ( $d = 5, p = 100$ ). Inferred network (green = TP, red = FP, orange = FN). Fully drawn lines are true ODE solutions, dashed lines are least squares estimates using inferred network.

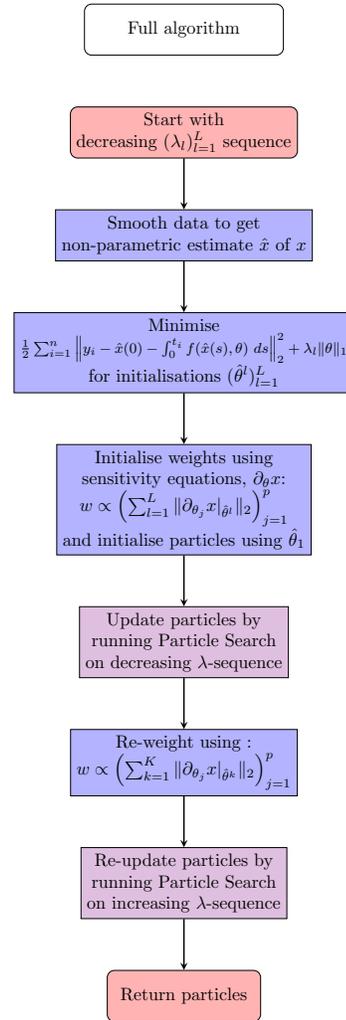


Fig. 4. Flowchart of full algorithm utilising the particle method.

- [4] J. Kennedy and R. Eberhart, *Particle Swarm Optimization*, Proceedings of IEEE International Conference on Neural Networks. IV. pp. 1942-1948
- [5] 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.
- [6] S. Yun and K-C Toh, *A coordinate gradient descent method for  $l_1$ -regularized convex minimization*, Computational Optimization and Ap-

- plications (2011), 48(2), 273-30.
- [7] H. Zou, *The Adaptive Lasso and Its Oracle Properties*, Journal of the American Statistical Association, 101:476, 1418-1429