

# Efficient Sampling by Marginalization of Scaling Parameters for ODE Models with Relative Data

Elba Raimúndez<sup>\*,\*\*,\*\*\*</sup> Jan Hasenauer<sup>\*,\*\*,\*\*\*</sup>

<sup>\*</sup> *Helmholtz Zentrum München-German Research Center for Environmental Health, Institute of Computational Biology, 85764 Neuherberg, Germany*

<sup>\*\*</sup> *Center for Mathematics, Technische Universität München, 85748 Garching, Germany*

<sup>\*\*\*</sup> *Faculty of Mathematics and Natural Sciences, University of Bonn, 53113 Bonn, Germany (e-mails: {elba.raimundez, jan.hasenauer}@uni-bonn.de)*

## Introduction

Quantitative mathematical models have become important tools for understanding and unraveling the mechanisms underlying biological signaling processing [Klipp et al., 2005]. However, the parameters of these models are in general unknown a priori and need to be inferred from experimental data [Raue et al., 2013].

Most measurement techniques, e.g. Western blotting and fluorescence microscopy, only provide relative information about the absolute molecular state. To establish a proper statistical link between experimental data and a computational model, scaling parameters and noise parameters need to be introduced [Loos et al., 2018] hence increasing the dimensionality of the estimation problem.

In previous studies, hierarchical optimisation approaches for parameter estimation [Weber et al., 2011, Loos et al., 2018, Schmiester et al., 2019] have been developed. These approaches exploit that the scaling and noise parameters can be computed analytically for a given set of model parameters. These approaches have shown a substantial benefit in the optimisation convergence for many optimisation methods and improved the conditioning of the optimisation problem. However, these concepts cannot be used for rigorous Bayesian uncertainty analysis.

In this study, we build on the idea of hierarchical optimisation and determine an analytical marginalization of the posterior with respect to the scaling and noise parameters. This yields a lower dimensional posterior which can more easily be explored using Markov Chain Monte Carlo (MCMC) sampling or related approaches.

## Parameter estimation problem

We consider the ordinary differential equation (ODE) model

$$\frac{dx(t, \theta)}{dt} = f(x(t, \theta), \theta), \quad x(t_0, \theta) = x_0(\theta)$$

$$y(t, \theta, s) = s \cdot h(x(t, \theta), \theta)$$

with state vector  $x$  and scalar observable  $y$ , depending in the vector field  $f$  and the observation function  $h$ . The parameter vector influencing the state variables is denoted by  $\theta$ , and the scalar scaling factor is denoted by  $s$ .

For this model, we consider measurement data  $D$  with normally distributed measurement noise with variance  $\sigma^2$ ,

$$\bar{y}_k = y(t_k, \theta, s) + \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, \sigma^2)$$

with indices  $k = 1, \dots, N$  for the time points. The kinetic parameters  $\theta$ , the scaling factors  $s$  and the noise variance  $\sigma^2$  are unknown and need to be estimated from the experimental data  $D$ .

Following Bayes' theorem, the posterior distribution is

$$p(\theta, s, \sigma^2 | D) = \frac{p(D | \theta, s, \sigma^2) p(\theta) p(s) p(\sigma^2)}{p(D)}.$$

A representative sample from  $p(\theta, s, \sigma^2 | D)$  can be obtained using MCMC sampling.

## Analytical evaluation of marginal likelihood

For a mechanistic understanding of the biological processes only the parameter vector  $\theta$  is required, but not the scaling and noise parameters. Therefore, we consider the marginalization of the posterior  $p(\theta, s, \sigma^2 | D)$  with respect to  $s$  and  $\sigma^2$ , yielding

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{p(D)}$$

with

$$p(D | \theta) = \int_0^\infty \int_{-\infty}^\infty p(D | \theta, s, \sigma^2) p(s) p(\sigma^2) ds d\sigma^2.$$

In this study, we consider the simplified case uniform priors,  $p(s) = p(\sigma^2) = 1$ . These priors are improper, however, for sampling this is not problematic as it holds that

$$p(\theta | D, s, \sigma^2) \propto p(D | \theta, s, \sigma^2) p(\theta) p(s) p(\sigma^2) \\ \propto p(D | \theta, s, \sigma^2) p(\theta).$$

The likelihood of  $D$  given  $\theta$  for  $h_k := h(x(t_k, \theta), \theta)$  is

$$p(D | \theta, s, \sigma^2) = \prod_{k=1}^N \mathcal{N}(\bar{y}_k | s \cdot h_k, \sigma^2) = \\ = \prod_{k=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{(\bar{y}_k - s \cdot h_k)^2}{\sigma^2}\right\}$$

yielding

$$p(D|\theta) = \int_0^\infty \frac{1}{(2\pi\sigma^2)^{N/2}} \underbrace{\int_{-\infty}^\infty \exp\left\{-\frac{1}{2} \sum_{k=1}^N \frac{(\bar{y}_k - s \cdot h_k)^2}{\sigma^2}\right\}}_{(*)} ds d\sigma^2. \quad (1)$$

The integral (\*) over the scaling parameter can be reformulated to

$$(*) = \int_{-\infty}^\infty \exp\left\{-s^2 \frac{1}{2\sigma^2} \sum_{k=1}^N h_k^2 + s \frac{1}{\sigma^2} \sum_{k=1}^N \bar{y}_k h_k - \frac{1}{2\sigma^2} \sum_{k=1}^N \bar{y}_k^2\right\} ds.$$

This Gaussian integral [Owen, 1980] takes the value

$$(*) = \sqrt{\frac{2\pi\sigma^2}{\sum_{k=1}^N h_k^2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{k=1}^N \bar{y}_k^2 - \frac{(\sum_{k=1}^N \bar{y}_k h_k)^2}{\sum_{k=1}^N h_k^2}\right)\right\}$$

Substitution of (\*) in (1) yields

$$p(D|\theta) = \frac{1}{\sqrt{\sum_{k=1}^N h_k^2}} \int_0^\infty \frac{1}{(2\pi\sigma^2)^{(N-1)/2}} \exp\left\{-\frac{w}{\sigma^2}\right\} d\sigma^2$$

with

$$w = \frac{1}{2} \left( \sum_{k=1}^N \bar{y}_k^2 \sum_{k=1}^N h_k^2 - \left( \sum_{k=1}^N \bar{y}_k h_k \right)^2 \right) \left( \sum_{k=1}^N h_k^2 \right)^{-1}$$

Using the transformation  $\sigma^2 = w/\gamma$  we finally get

$$p(D|\theta) = \frac{w^{(N-3)/2}}{(2\pi)^{(N-1)/2}} \frac{1}{\sqrt{\sum_{k=1}^N h_k^2}} \times \int_0^\infty \gamma^{(N-3)/2-1} \exp\{-\gamma\} d\gamma.$$

As it holds that

$$\int_0^\infty t^{b-1} \exp\{-t\} dt = \Gamma(b)$$

we obtain

$$p(D|\theta) = \frac{w^{(N-3)/2}}{(2\pi)^{(N-1)/2}} \frac{1}{\sqrt{\sum_{k=1}^N h_k^2}} \Gamma\left(\frac{N-3}{2}\right).$$

The substitution of  $w$  yields

$$p(D|\theta) = \frac{1}{2\pi^{(N-1)/2}} \times \left( \sum_{k=1}^N \bar{y}_k^2 \sum_{k=1}^N h_k^2 - \left( \sum_{k=1}^N \bar{y}_k h_k \right)^2 \right)^{-(N-3)/2} \times \left( \sum_{k=1}^N h_k^2 \right)^{(N-4)/2} \Gamma\left(\frac{N-3}{2}\right). \quad (2)$$

### Sampling of marginal posterior distribution

The marginal likelihood (2) provides the likelihood of the data  $D$  for a parameter vector  $\theta$ , given the uncertainty

in the unknown scaling parameter  $s$  and the unknown noise variance  $\sigma^2$ . Hence, the analytical marginalization enables the estimation of  $\theta$  without knowledge about  $s$  and  $\sigma^2$ . The posterior distribution of  $\theta$  can be sampled using, e.g. adaptive Metropolis [Haario et al., 2001] and parallel tempering [Lacki and Miasojedow, 2015]. These methods will simply use the analytical marginal likelihood (2) to sample the marginal posterior  $p(\theta|D)$  – instead of the full posterior  $p(\theta, s, \sigma^2|D)$ .

### Discussion

We presented an analytical marginalization of the likelihood function for estimating the parameters of ODE models from relative data. As the marginalization enables parameter sampling in a reduced space, we expect improved mixing and convergence compared to standard approach. Initial results on increasingly complex models in an established benchmark collection [Hass et al., 2019] confirm these results.

We presented the approach for ODE models, but conceptually it is also applicable for other modelling approaches. Furthermore, as many experimental techniques provide relative measurements, the approach is applicable to a broad spectrum of problems in systems biology, engineering and physics. In a next step, we plan to extend it to proper priors, e.g. normal distributions for  $p(s)$  and  $p(\sigma^2)$ .

### REFERENCES

- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.
- Hass, H., Loos, C., Raimúndez-Álvarez, E., Timmer, J., Hasenauer, J., and Kreutz, C. (2019). Benchmark problems for dynamic modeling of intracellular processes. *Bioinf.*, btz020.
- Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2005). *Systems biology in practice*. Wiley-VCH, Weinheim.
- Lacki, M.K. and Miasojedow, B. (2015). State-dependent swap strategies and automatic reduction of number of temperatures in adaptive parallel tempering algorithm. *Stat. Comput.*, 26(5), 951–964.
- Loos, C., Krause, S., and Hasenauer, J. (2018). Hierarchical optimization for the efficient parametrization of ODE models. *Bioinf.*, 34(24), 4266–4273.
- Owen, D.B. (1980). A table of normal integrals: A table. *Communications in Statistics-Simulation and Computation*, 9(4), 389–419.
- Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B.D., Theis, F.J., Klingmüller, U., and Timmer, J. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, 8(9), e74335.
- Schmiester, L., Schälte, Y., Fröhlich, F., Hasenauer, J., and Weindl, D. (2019). Efficient parameterization of large-scale dynamic models based on relative measurements. *Bioinf.* btz581.
- Weber, P., Hasenauer, J., Allgöwer, F., and Radde, N. (2011). Parameter estimation and identifiability of biological networks using relative data. In *Proc. of the 18th IFAC World Congress*, volume 18, 11648–11653.