# Inference in nonlinear differential equations

Mu Niu[1], Maurizio Filippone[2], Dirk Husmeier[1], Simon Rogers[2]

[1] School of Mathematics and Statistics, University of Glasgow, Glasgow UK
[2] School of Computing Science, University of Glasgow, Glasgow UK

E-mail for correspondence: mu.niu@glasgow.ac.uk

**Abstract:** Parameter inference in mechanistic models of coupled differential equations is a challenging problem. We propose a new method using kernel ridge regression in Reproducing Kernel Hilbert Spaces (RKHS). A three-step gradient matching algorithm is developed and applied to a realistic biochemical model.

**Keywords:** RKHS; nonlinear ordinary differential equation; gradient matching.

## 1 Introduction

Many processes in science and engineering can be described by dynamical systems models based on nonlinear ordinary differential equations (NODEs). However, the parameters of the NODEs are often unknown and not directly measurable. Direct inference requires a computationally expensive numerical integration of the NODEs every time the parameters are adapted. For that reason, approximate methods based on gradient matching have recently gained much attention; see e.g. Dondelinger et al. (2013) and Heinonen et al.(2014). The purpose of the present article is to try a new variant of this approach based on reproducing kernel Hilbert spaces (RKHS). We consider systems governed by first order multivariate NODEs:

$$\dot{x} = \frac{dx}{dt} = f\big(x(t), \theta\big), \tag{1}$$

with initial value $x(t_1) = x_1$, where $x(t)$ is an $r$ dimensional vector of state variables ($x(t) \in \mathbb{R}^r$). We observe the states of the system at $n$ time points $(y_1, ..., y_n)$ and assume that the observations consist of the states corrupted by Normal additive iid noise $y_i = x(t_i) + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$ iid. In this work, a framework of penalized regression for vector-valued functions in RKHS is used to make predictions of the system state. A three step gradient matching approach is developed to learn the NODEs parameters:

---

This paper was published as a part of the proceedings of the 30th International Workshop on Statistical Modelling, Johannes Kepler Universität Linz, 6–10 July 2015. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

1. Learn a smooth function $g$ over the observations using an RKHS approximation with observation $(y_1, ..., y_n)$ at time points $(t_1, ..., t_n)$.
2. Given $g$, learn the NODEs parameters $\theta$ by gradient matching to minimize the difference between $\dot{g}(t)$ and $f(g(t), \theta)$.
3. Re-estimate $g$ using the optimised $\theta$ and observed trajectory $y$. Given re-estimated $g$, optimise $\theta$ as in step 2.

## 2 Methodology

For each state variable $x_s$ indexed by $s = 1, ..., r$, a positive definite kernel $k_s : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is used to define the Hilbert space $\mathcal{G}_s$ of the smoother function $g_s$. The squared exponential kernel (Gaussian kernel) is used in this study with lengthscale parameter $l$: $k(t_k, t_i) = \exp\big(-l^{-2}(t_k - t_i)^2\big)$. In step one, kernel ridge regression is used to learn $g_s$, the smoother of the $s$th state with the following loss function:

$$\mathcal{L}_1(l_s) = \sum_{i=1}^{n}(g_s(t_i) - y_{s_i})^2 + ||g_s||^2 \tag{2}$$

$$g_s(t_i) = \sum_{k=1}^{n} b_{s_k} k_s(t_k, t_i) \tag{3}$$

$$\overline{b}_s = (K_s + \lambda_s I)^{-1}\overline{y}_s, \qquad ||g_s||^2 = \lambda_s \overline{b}_s^T K_s \overline{b}_s, \tag{4}$$

where $k_s(., t_i)$ is the $i$th basis function and $K_s$ is the Gram kernel matrix. $\overline{y}_s$ is the vector of observations for the $s$th state, $||g_s||$ is the norm of $\mathcal{G}_s$ and the regularisation factor $\lambda_s$ is estimated using leave-one-out cross validation. $I$ is the $n \times n$ identity matrix. Different states are assumed to have different lengthscales, $l_s$, and these are estimated independently through a gradient-based quasi-Newton optimisation routine. With the optimised $l_s$, the state estimation $g_s(t_j)$ of $x_s(t_j)$ is made on a uniform grid of $m$ time points indexed by $j = 1, ..., m$. The gradient of the smoother is:

$$\dot{g}_s(t_j) = \sum_{i=1}^{n} b_{s_i} \frac{dk(t_i, t_j)}{dt_i}. \tag{5}$$

In step 2, the NODE parameters are estimated by gradient matching for all states with the loss function:

$$\mathcal{L}_2(\theta) = \sum_{s=1}^{r}\sum_{j=1}^{m}\Big(\dot{g}_s(t_j) - f\big(g_s(t_j), \theta\big)\Big)^2, \tag{6}$$

where $f\big(g_s(t_j), \theta\big)$ is the target gradient generated by feeding the output of the $s$th smoother, $g_s(t)$, into the NODEs system. $\theta$ is optimised using a gradient-based quasi-Newton routine.
Finally, in step 3, we use the estimated $\theta$ from step 2 to construct the NODE system. The lengthscale vector $\overline{l}$ for all state variables is then re-estimated

using the full loss function.

$$\mathcal{L}_3(\bar{l}) = \sum_{s=1}^{r}\Big(\sum_{i=1}^{n}(g_s(t_i)-y_{s_i})^2 + ||g_s||^2 + \sum_{j=1}^{m}(\dot{g}_s(t_j) - f\big(g_s(t_j),\theta\big)\big)^2\Big) \quad (7)$$

The updated smoother $\tilde{g}(t)$ is calculated using $\bar{l}$. The refined $\theta$ can then be optimised iteratively by repeating the second step.

## 3 Application: Calcium model

The calcium model (Peifer and Timmer, 2007) represents the oscillations of calcium signaling in eukaryotic cells via a dynamic system with states corresponding to the concentrations of free calcium in cytoplasm $C_{ac}$ and endoplasmic reticulum $C_{ar}$, as well as active $G_\alpha$. and phospholipase-C, $P_c$.

$$\begin{aligned}
\frac{dG_\alpha}{dt} &= k_1 + k_2 G_\alpha - k_3 P_c R_1(G_\alpha) - k_4 C_{ac} R_2(G_\alpha) \\
\frac{dP_c}{dt} &= k_5 G_\alpha - k_6 R_3(P_c) \\
\frac{dC_{ac}}{dt} &= k_7 P_c C_{ac} R_4(C_{ar}) + k_8 P_c + k_9 G_\alpha - k_{10} R_5(C_{ac}) - k_{11} R_6(C_{ac}) \\
\frac{dC_{ar}}{dt} &= -k_7 P_c C_{ac} R_4(C_{ar}) + k_{11} R_6(C_{ac}) \qquad (8)
\end{aligned}$$

where $R_i(x) = \frac{x}{x+km_i}$. We followed Oates et al.(2014) and fixed the $km_i$ parameters, leaving $k_{1:11}$ to be inferred. Three scenarios were tested: the noise-free case, and adding iid Gaussian noise with signal-to-noise ratio $SNR = 50db$ and $SNR = 10db$ to the numerical solution of the model. We sampled $n = 100$ ($t_1 = 0, t_{100} = 20$) regularly spaced observations, and made state predictions on $m = 200$ grid points. For $10db$ and $50db$ noise, 100 independent data sets were generated.

The mean NODEs parameter estimates with error bars are shown in Figure 1. We see that when no noise is added, our estimates closely agree with the true values. The parameter estimates at $SNR = 50db$ noise are similar to the noise-free case when the parameters are smaller than 10, but drift away from the true value for parameters larger than 10. For the highest noise level, $SNR = 10db$, the NODEs parameter estimates become worse, although we still have 4 reliable estimates out of 11 parameters. For the $SNR = 50db$ case, the length scale parameters $l_{1:4}$ for the 4 states at the optimum after step 2 are $[0.06, 0.18, 0.01, 0.05]$. After step 3, these change to [0.06,0.81,0.22,0.79], suggesting that a regularising effect from the NODEs exists.

## 4 Conclusion

We have described an RKHS based gradient matching approach for parameter inference in a system of NODEs. In low noise scenarios we obtain

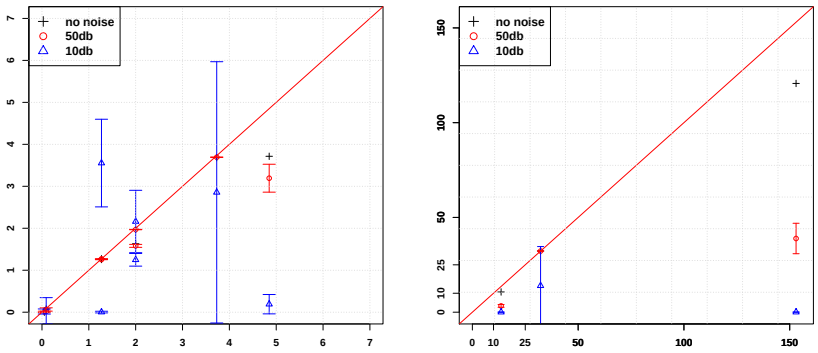


FIGURE 1: Scatter plot of true versus estimated NODEs parameters in three noise scenarios. No noise (black crosses), $SNR = 50db$ (red circles) and $SNR = 10db$ (blue triangles) noise. Left and right plots show parameters with true values $< 10$ and $> 10$ respectively.

good parameter estimates for a realistic biochemical application, and we have quantified the deterioration resulting from increased noise levels. We have demonstrated that in the proposed three-step algorithm, the length scale parameters of the kernels are regularised by the gradient matching step. A potential explanation for the poor performance in the high-noise regime is related to the choice of kernel. In our future work, we will explore the effect of using non-stationary kernels (like the MLP kernel) and less smooth kernels (like the Matérn class kernels) as alternatives to the squared exponential kernel.

**Acknowledgments:** This project is funded by the EPSRC (EP/L020319/1).

## References

Dondelinger, F. et al. (2013) ODE parameter inference using adaptive gradient matching with Gaussian processes. *JMLR W&CP* **31**: 216-228.

Heinonen, M., and d'Alch-Buc, F. (2014). Learning nonparametric differential equations with operator-valued kernels and gradient matching. *arXiv preprint arXiv:1411.5172*.

Oates, C.J., et al. (2014). Causal network inference using biochemical kinetics. *Bioinformatics*,**30**(17), i468-i474.

Peifer, M. and Timmer, J. (2007). Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. *Systems Biology, IET*,**1.2.**, 78 – 88.