Discussion of “Parameter Estimation for Differential Equations: A Generalized Smoothing Approach”

By SYLVAIN SARDY*

The backbone of the proposed methodology is the expansion representation of output functions so that both $x_i(t) = \sum_{k=1}^{K_i} c_{ik}\phi_i(t)$ and its derivative $\dot{x}_i(t) = \sum_{k=1}^{K_i} c_{ik}\dot{\phi}_i(t)$ are a linear form of the same coefficients $\mathbf{c}_i$. Besides solving the non-trivial optimization, providing variance estimation is also an achievement. Their substantial work is the source of many research directions for statisticians, like nonparametric estimation.

The authors essentially solve the least squares problem for systems of differential equations by letting λ_i get large in (13). At the limit, no regularization is performed: they solve the constrained problem, as in linear regression one could solve $\min_{\boldsymbol{\theta}} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2$ by successively solving $\min_{\boldsymbol{\theta},\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{x} - X\boldsymbol{\theta}\|_2^2$ and letting $\lambda \to \infty$. This observation leads to two points. First the constrained optimization could be solved efficiently by handling constraints directly. Second if the true parametric equations are not completely known, the practice is to do model selection. Take the FitzHugh-Nagumo equations (2) for instance, one could start with the richer model

$$\begin{aligned} \dot{V} &= c\left(V - \frac{V^3}{3} + dR + eR^2\right) \\ \dot{R} &= -\frac{1}{c}\left(fV + g\log(V) - a + bR + hR^2\right) \end{aligned}$$

and estimate a sparse vector of coefficients $\boldsymbol{\theta} = (a, b, c, d, e, f, g, h)$ while satisfying the constraints imposed by the differential equations. A possible model selection strategy consists in solving a lasso-type ℓ_1-penalized least squares. The convex ℓ_1 penalty on $\boldsymbol{\theta}$ may also have the advantage of getting rid of some of the ripples of Figure 2. Solving the constrained ℓ_1 penalized least squares is a worthy challenge to achieve model selection for systems of nonlinear differential equations. Finally, increasing the dimension of $\boldsymbol{\theta}$ with the sample size, nonparametric estimation becomes possible.

The number of terms K_i used in each spline-expansion relative to the number Q_i of collocation points is important. We illustrate with a toy differential equation: $\dot{x}(t) = f(t, x; \theta) = \theta x(t)$ defined on $[-2, 2]$ with $\theta =$

*Address for correspondence: Université de Genève, Section de mathématiques, 2-4 rue du Lièvre Case postale 64, 1211 Genève 4, Suisse. Email: Sylvain.Sardy@epfl.ch

1.3. The data consists in $N = 500$ equispaced measurements $\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, I)$. We solve

$$\min_{\theta,\mathbf{c}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \quad \text{s.t.} \quad \begin{cases} \mathbf{L}(\mathbf{c}, \theta) := \dot{B}\mathbf{c} - \theta B\mathbf{c} = \mathbf{0} \\ \mathbf{x} = X\mathbf{c} \end{cases},$$

where X is the $N \times K$ spline matrix ($K = 250$), and B and $\dot{B}$ are the $Q \times K$ matrices of splines and their derivatives. For the residuals $\mathbf{L}(\mathbf{c}, \theta) = \mathbf{0}$ to have a solution other than $\mathbf{c} = \mathbf{0}$ for all θ, the kernel of $\dot{B} - \theta B$ must be different from $\{\mathbf{0}\}$. A sufficient condition is $Q < K$. Here the correct choice seems to be $Q = K - 1$ to have $\dim(\mathrm{Ker}(\dot{B} - \theta B)) = 1$ and to fit the solution $x(t; \theta) = \exp(\theta t)$. Choosing a smaller value of Q has adverse effects. For each choice of $Q \in \{K-1, K-2, K-50\}$ we estimate 500 times θ and $\mathbf{c}$. Looking at Figure 1, we see that both bias and variance increase with $K - Q$ and that the estimation of $\dot{x}(t)$ becomes bad when $Q < K - 1$. This Monte

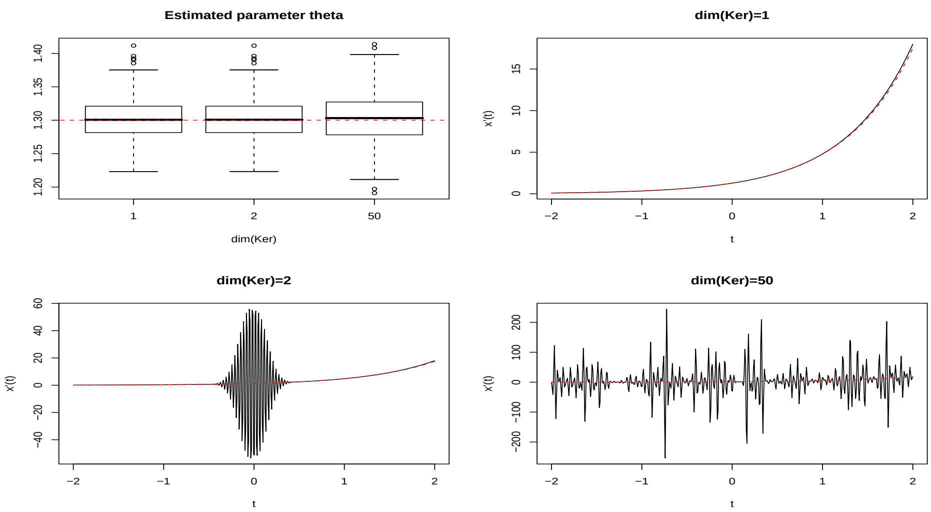

Figure 1: Monte Carlo simulation for $\theta = 1.3$ with Gaussian equispaced samples of size 500. Top left: boxplots of $\hat{\theta}$ for $Q \in \{K-1, K-2, K-50\}$. One typical estimated derivative of $\dot{x}(t)$ for the respective values of Q.

Carlo experiment shows that the choice of K_i and Q_i calls for particular attention.