On parameter estimation for delay models with discontinuous right-hand sides

M. Ashyraliyev

MAS-E0908

Centrum Wiskunde & Informatica

On parameter estimation for delay models with discontinuous right-hand sides

ABSTRACT
We study delay models with discontinuous right-hand side. Lack of smoothness in the solutions of such problems may have serious consequences for parameters estimation using gradient-based approaches. Additionally, it may cause ambiguities in the parameter determinability analysis applied on the parameter estimates. In order to overcome these difficulties, we suggest a standard regularization technique to make the model continuous. We prove the convergence of the solution of the regularized model to the solution of the original problem. As a consequence of that, parameter estimates inferred from the regularized model converge to the corresponding estimates of the original problem. We support our findings with numerical illustrations for simple test problems.

2000 Mathematics Subject Classification: 65K10, 65Q05
Keywords and Phrases: Delay Differential Equations, parameter estimation, regularization

On Parameter Estimation for Delay Models with Discontinuous Right-Hand Sides

M. Ashyraliyev
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Abstract

We study delay models with discontinuous right-hand side. Lack of smoothness in the solutions of such problems may have serious consequences for parameter estimation using gradient-based approaches. Additionally, it may cause ambiguities in the parameter determinability analysis applied on the parameter estimates. In order to overcome these difficulties, we suggest a standard regularization technique to make the model continuous. We prove the convergence of the solution of the regularized model to the solution of the original problem. As a consequence of that, parameter estimates inferred from the regularized model converge to the corresponding estimates of the original problem. We support our findings with numerical illustrations for simple test problems.

1 Introduction

Many real-life processes can be modeled by non-linear Ordinary Differential Equations (ODEs) or Partial Differential Equations (PDEs). There is a growing number of applications which use Delay Differential Equations (DDEs). For instance, delay models are widely used in biosciences [1]-[2]. Models with DDEs are usually more complicated than corresponding ODE or PDE models, both with regard to theoretical analysis and to numerical simulation. However, for many problems it has been shown that DDE models provide a better description of reality, as well as a better understanding of the process under consideration. The increasing number of such applications stimulates us to study the common numerical challenges one may encounter while using DDE models. It is well-known that the solutions of DDEs suffer from lack of smoothness [3]. This can be a property of ODE or PDE solutions as well. However, for ODE and PDE cases it is rather an exception, while for DDEs it is more like a rule.

Similar to ODE and PDE models, DDE models may have a number of unknown parameters among which there can be unknown time lags. Sometimes, it is feasible to determine the missing parameters experimentally, but in most cases this is difficult or even impossible. However, one can usually measure other quantities involved in the model. Unknown model parameters can then be found by parameter estimation techniques based on fitting the model solution to the measured data. Lack of smoothness for the solutions of DDEs can have serious consequences for parameter estimation [4]. Gradient-based optimization methods are based on sufficient smoothness of the solution. So, once the solution of a DDE or its derivatives

have discontinuities, gradient-based methods may fail to find parameter estimates. Despite the discontinuities parameter inference with gradient-based methods may succeed in some cases, but in general for such models one has to use gradient-free optimization techniques (such as direct search) or methods based on generalized gradients [5].

Once the parameters are found by means of optimization, it is important to assess the quality of the parameter estimates, e.g., by computing confidence intervals. This is especially important if one wants to deduce for the real-life process under consideration quantitative or qualitative conclusions based on parameter estimates. Such conclusions may greatly depend on how reliable the estimates are. One also might be interested in computing the correlations between parameters. The computation of confidence intervals and correlations is based on sufficient smoothness of the solution at the neighborhood of parameter estimates. Therefore, for DDE models, even if one succeeds with parameter estimation, still statistical analysis may be ambiguous in some cases.

In this report we will focus our attention on a specific type of non-linear DDE models with application in developmental biology. These models do have a discontinuous right-hand side such that the derivative with respect to parameters is discontinuous. We will discuss how the discontinuities arise in the gradients of the solution. We will demonstrate how these discontinuities may affect the estimation of parameters and the statistical analysis. There is a finite number of such discontinuity points. Unlike in the case of Neutral Delay Differential Equations (NDDEs), they do not propagate in time because it is well-known that the solution of DDEs smooths out [3]. As a natural remedy, we use the regularization technique to make the right-hand side of the DDEs continuous. Since the time lags are constant, this can be done in a straightforward manner. To apply such a technique in practice, one of course has to ask the following questions. How large is the difference between the solutions of the original and the regularized model? How large is the difference between their gradients? How much differ the parameter estimates inferred from original and regularized models? How does regularization affect the statistical analysis? Finally, it is important to know how all these issues depend on the type of regularization used. We shall answer these questions in this report by means of theoretical proofs and numerical illustrations for our test problems. Note that our approach is similar to the regularization method used in [8] for NDDEs having discontinuous solutions. The main significance of our results is that we are able to derive the rate of convergence of the solution of the regularized problem to the solution of the original problem. Namely, we will show that the solution of the regularized problem, which can be also considered as an approximate solution, has second-order convergence to the solution of the original problem almost everywhere in time and first-order convergence only in the vicinity of discontinuity points. Our convergence results do not depend explicitly on the way the model is regularized. Therefore, with a specific choice of regularization we achieve not only the continuity of the right-hand side of the model but also its continuous differentiability gaining extra smoothness in the solution of the regularized model. Finally, our results imply that the parameter values estimated from the regularized model converge to corresponding estimates of the original problem.

The report is organized as follows. In Section 2 we introduce the notations, describe the problem, outline the main difficulties caused by discontinuities, and define the regularized problem to overcome them. In Section 3 we prove the convergence theorem for the solution of the regularized model. Section 4 supports our findings with numerical illustrations. We conclude the note with a discussion in Section 5.

2 Notations and problem description

We consider DDEs of the form

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{y}}{\mathrm{d}t}(t,\theta) &= \mathrm{H}(t-\tau)\ \mathbf{f}(t,\mathbf{y}(t-\tau,\theta),\lambda) - \mathrm{A}(\lambda)\ \mathbf{y}(t,\theta), \quad 0 < t \le T, \\ \mathbf{y}(t,\theta) &= \Psi(t), \qquad\qquad\qquad\qquad\qquad\qquad\quad t \le 0, \end{cases} \tag{2.1}$$

where $\mathbf{y}$ is an n-dimensional state vector, $\mathbf{f}$ is a given nonlinear vector function, continuous and sufficiently differentiable with respect to all arguments, τ is an unknown time lag [1)], λ is an $(m-1)$-dimensional (unknown) parameter vector, $\theta = \{\lambda, \tau\}$ [2)] is a vector of size m, A is a real-valued symmetric matrix, the initial function $\Psi(t)$ is continuous, and H is the Heaviside function

$$\mathrm{H}(t) = \begin{cases} 0, & t < 0, \\ 1, & t \ge 0. \end{cases} \tag{2.2}$$

For simplicity of presentation, without loss of generality, we assume that the initial function Ψ does not depend on the parameter vector θ. This type of DDEs finds an application in the modelling of regulatory networks. The non-linear part in the right-hand side of (2.1) models protein synthesis, while the linear part includes protein decay and the spatial discretization of protein diffusion [7].

Note that the solution of (2.1) does not depend on $\Psi(t)$ $(t < 0)$. In fact, only $\Psi(0)$ is needed for the mathematical definition of the model. However, as we shall see in the next section, if $\Psi(t)$ $(t < 0)$ is available then using that knowledge improves the convergence of the solution of the regularized model to the solution of (2.1).

Let us assume that for fitting (2.1) there are N measurements available. Each measurement, which we denote by $\tilde{y}_i$, is specified by the time t_i when the c_i-th component of the state vector $\mathbf{y}$ is measured. The corresponding model value obtained from (2.1) is denoted by $y_{c_i}(t_i,\theta)$. With the chosen notation, two measurements $y_{c_i}(t_i,\theta)$ and $y_{c_j}(t_j,\theta)$ $(i \ne j)$ can indicate one of the following:

- two different components at two different time points (when $c_i \ne c_j, t_i \ne t_j$);
- the same component at two different time points (when $c_i = c_j, t_i \ne t_j$);
- two different components at the same time point (when $c_i \ne c_j, t_i = t_j$).

We denote the vector of weighted discrepancies between the model outputs and the measured values by $\mathrm{Y}(\theta)$. Then the least squares estimate $\hat{\theta}$ of the parameters is the value of θ that minimizes the sum of squares

$$S(\theta) = \sum_{i=1}^{N} w_i^2 (y_{c_i}(t_i,\theta) - \tilde{y}_i)^2 = \mathrm{Y}^T(\theta)\mathrm{Y}(\theta), \tag{2.3}$$

where w_i are positive weights. If the measurement errors in $\tilde{y}_i$ are independent of each other and normally distributed with zero mean and standard deviations σ_i, then the minimization of (2.3), with the weights w_i being inversely proportional to the corresponding standard deviations σ_i, yields the Maximum Likelihood Estimate (MLE) [9].

[1)] For ease of presentation, we assume that there is only one time lag. The presented material is extendable to the case with more than one time lag.

[2)] We distinguish between the parameters τ and λ because they feature in a different way in the analysis. Note that the right-hand side of (2.1) is continuous in λ and discontinuous in τ.

Remark 2.1. From the continuity of the solution of (2.1) it follows that the objective function (2.3) is continuous in the parameters contained in θ.

The gradient-based optimization methods require the objective function $S(\theta)$ to be continuously differentiable with respect to θ. Differentiation of (2.3) with respect to component θ_j, $j = 1, 2, \ldots, m$, gives

$$\frac{\partial S}{\partial \theta_j}(\theta) = 2 \sum_{i=1}^{N} w_i^2 (y_{c_i}(t_i, \theta) - \tilde{y}_i) \frac{\partial y_{c_i}}{\partial \theta_j}(t_i, \theta) = 2 \, \mathrm{Y}^T(\theta) \frac{\partial \mathrm{Y}}{\partial \theta_j}(\theta). \tag{2.4}$$

Obviously, discontinuities in $\frac{\partial S}{\partial \theta_j}$ can arise only from discontinuities in $\frac{\partial \mathbf{y}}{\partial \theta_j}$. We shall investigate here under which circumstances this may happen. The derivatives $\frac{\partial \mathbf{y}}{\partial \theta_j}$ can be found by solving the so-called variational equations. First, we denote

$$\mathbf{f}_{\mathbf{y}} = \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \mathbf{y}, \theta), \quad \mathbf{f}_{\lambda} = \frac{\partial \mathbf{f}}{\partial \lambda}(t, \mathbf{y}, \theta), \quad \mathbf{f}_{\tau} = \frac{\partial \mathbf{f}}{\partial \tau}(t, \mathbf{y}, \theta). \tag{2.5}$$

Differentiating (2.1) with respect to λ_i, $i = 1, 2, \ldots, m-1$, we have

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} \dfrac{\partial \mathbf{y}}{\partial \lambda_i}(t, \theta) &= \mathrm{H}(t-\tau) \left[\mathbf{f}_{\mathbf{y}} \, \dfrac{\partial \mathbf{y}}{\partial \lambda_i}(t-\tau, \theta) + \mathbf{f}_{\lambda_i} \right] \\ & \quad - \dfrac{\partial \mathrm{A}}{\partial \lambda_i}(\lambda) \, \mathbf{y}(t, \theta) - \mathrm{A}(\lambda) \, \dfrac{\partial \mathbf{y}}{\partial \lambda_i}(t, \theta), \quad 0 < t \le T, \\ \dfrac{\partial \mathbf{y}}{\partial \lambda_i}(t, \theta) &= 0, \qquad\qquad t \le 0. \end{cases} \tag{2.6}$$

Since the right-hand side of (2.1) is continuously differentiable with respect to λ_i, the derivation of (2.6) is valid.

Remark 2.2. From the continuity of the solution of (2.6) it follows that the gradient $\frac{\partial S}{\partial \lambda_i}$ is continuous in the parameters θ_j, $j = 1, 2, \ldots, m$.

Differentiating (2.1) with respect to τ, we obtain

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} \dfrac{\partial \mathbf{y}}{\partial \tau}(t, \theta) &= -\delta(t-\tau) \, \mathbf{f} - \mathrm{A}(\lambda) \, \dfrac{\partial \mathbf{y}}{\partial \tau}(t, \theta) \\ & \quad + \mathrm{H}(t-\tau) \, \mathbf{f}_{\mathbf{y}} \left[\dfrac{\partial \mathbf{y}}{\partial \tau}(t-\tau, \theta) - \dfrac{\mathrm{d}\mathbf{y}}{\mathrm{d}t}(t-\tau, \theta) \right], \quad 0 < t \le T, \\ \dfrac{\partial \mathbf{y}}{\partial \tau}(t, \theta) &= 0, \qquad\qquad t \le 0, \end{cases} \tag{2.7}$$

where δ is the Dirac delta function. Note that (2.7) is defined at $t = \tau$ only in the generalized sense, i.e., the time derivative of the solution does not have to exist at that point. For example, the problem $\frac{\mathrm{d}v}{\mathrm{d}t} = \delta(t - t_0)$, $v(0) = 0$, defined in the generalized sense, has as solution $v(t) = H(t - t_0)$.

The solution of (2.7) is discontinuous at $t = \tau$. Although (2.7) is a system of NDDEs, the right-hand side of it does not depend on the time derivative of $\frac{\partial \mathbf{y}}{\partial \tau}$ but does depend on the time derivative of $\mathbf{y}$, the solution of (2.1), which smooths out in time. Thereby, the discontinuity in $\frac{\partial \mathbf{y}}{\partial \tau}$ at $t = \tau$ does not propagate in time which is a common property in general for solution of NDDEs.

Remark 2.3. The discontinuity in $\frac{\partial \mathbf{y}}{\partial \tau}$ at $t = \tau$ enters in $\frac{\partial S}{\partial \tau}$ if this point coincides with one of the data points, i.e., if $\tau = t_i$ for some $i = 1, \ldots, N$.

Example 2.4. Consider the simple problem

$$\begin{cases} \dfrac{\mathrm{d}y}{\mathrm{d}t}(t,\tau) = \mathrm{H}(t-\tau)y(t-\tau,\tau), & 0 < t \le 3\tau, \\ y(t,\tau) = 1, & t \le 0. \end{cases}$$

Its solution is

$$y(t,\tau) = \begin{cases} 1, & 0 \le t < \tau, \\ t - \tau + 1, & \tau \le t < 2\tau, \\ \dfrac{(t-2\tau+1)^2}{2} + \tau + \dfrac{1}{2}, & 2\tau \le t \le 3\tau. \end{cases}$$

Therefore,

$$\frac{\partial y}{\partial \tau}(t,\tau) = \begin{cases} 0, & 0 \le t < \tau, \\ -1, & \tau \le t < 2\tau, \\ -2t + 4\tau - 1, & 2\tau \le t \le 3\tau. \end{cases}$$

For this example y is continuous everywhere, while $\frac{\partial y}{\partial \tau}$ is discontinuous at $t = \tau$.

Discontinuities in $\frac{\partial y}{\partial \theta}$ may cause ambiguity for statistical analysis of parameter estimates, such as computing confidence intervals and correlations between parameters. Assume that $\hat{\theta}$ has been found by means of some optimization method. Dependent and independent confidence intervals for each estimate $\hat{\theta}_i$ $(i = 1, 2, \ldots, m)$ are given by

$$\left[\hat{\theta}_i - \triangle_i^D,\ \hat{\theta}_i + \triangle_i^D\right], \qquad \triangle_i^D = \sqrt{\frac{m}{N-m}\frac{S(\hat{\theta})\ F_\alpha(m, N-m)}{\left(J^T(\hat{\theta})\ J(\hat{\theta})\right)_{ii}}} \tag{2.8}$$

and

$$\left[\hat{\theta}_i - \triangle_i^I,\ \hat{\theta}_i + \triangle_i^I\right], \qquad \triangle_i^I = \sqrt{\frac{m}{N-m}\ S(\hat{\theta})\ F_\alpha(m, N-m)\left(\left(J^T(\hat{\theta})\ J(\hat{\theta})\right)^{-1}\right)_{ii}}, \tag{2.9}$$

respectively. Here $J = \frac{\partial \mathrm{Y}}{\partial \theta}$ is the so-called sensitivity matrix and $F_\alpha(m, N-m)$ is the upper α part of Fisher's distribution with m and $N - m$ degrees of freedom. The correlation coefficient between $\hat{\theta}_i$ and $\hat{\theta}_j$ is given by:

$$\rho_{ij} = \frac{B_{ij}}{\sqrt{B_{ii}B_{jj}}}, \tag{2.10}$$

where $B(\hat{\theta}) = \left(J^T(\hat{\theta})\ J(\hat{\theta})\right)^{-1}$. See [6] and references therein for explanations of these statistical quantities.

The computation of confidence intervals and correlation coefficients both are based on $J(\hat{\theta})$. If the discontinuities in $\frac{\partial \mathbf{y}}{\partial \theta}$ enter into J and the estimate itself is a discontinuity point, then (2.8)-(2.10) are not valid and one cannot deduce any conclusions.

So far, we have described two major difficulties if one uses model (2.1). Both of them are related to discontinuities in $\frac{\partial \mathbf{y}}{\partial \theta}$. These discontinuities stem from the discontinuous right-hand side of the model. To overcome these problems, we consider a regularized analogy of (2.1) given by:

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{z}}{\mathrm{d}t}(t,\theta) &= \mathrm{H}_\epsilon(t-\tau)\ \mathbf{f}(t,\mathbf{z}(t-\tau,\theta),\lambda) - \mathrm{A}(\lambda)\ \mathbf{z}(t,\theta), & 0 < t \le T, \\ \mathbf{z}(t,\theta) &= \Psi(t), & t \le 0, \end{cases} \tag{2.11}$$

where H_ϵ is a regularization of the Heaviside function (2.2) on the interval $[-\epsilon,\epsilon]$, with $0 < \epsilon \ll 1$. For instance, the regularization by linear and cubic polynomials are given by:

$$\mathrm{H}_\epsilon(t) = \begin{cases} 0, & t \le -\epsilon, \\ \dfrac{t}{2\epsilon} + \dfrac{1}{2}, & -\epsilon < t < \epsilon, \\ 1, & t \ge \epsilon, \end{cases} \tag{2.12}$$

and

$$\mathrm{H}_\epsilon(t) = \begin{cases} 0, & t \le -\epsilon, \\ -\dfrac{t^3}{4\epsilon^3} + \dfrac{3t}{4\epsilon} + \dfrac{1}{2}, & -\epsilon < t < \epsilon, \\ 1, & t \ge \epsilon, \end{cases} \tag{2.13}$$

respectively. As we shall see in Section 3, under certain assumptions the solution of (2.11) converges to the solution of the original problem (2.1) as $\epsilon \to 0$. None of those assumptions nor the convergence rate depend on the explicit form of H_ϵ. In fact, any continuous, monotonous and sufficiently differentiable on $(-\epsilon,\epsilon)$ function such that $\mathrm{H}_\epsilon(t) = 0,\ t \le -\epsilon$ and $\mathrm{H}_\epsilon(t) = 1,\ t \ge \epsilon$ is a suitable regularization of the Heaviside function (2.2). Any realization of $\mathrm{H}_\epsilon(t)$ has the following properties:

- Using $\mathrm{H}_\epsilon(-\epsilon) = 0,\ \mathrm{H}_\epsilon(\epsilon) = 1$, and Taylor expansion, we have $\mathrm{H}_\epsilon(0) = \frac{1}{2} + O(\epsilon^2)$;
- Monotonicity of $\mathrm{H}_\epsilon(t)$ implies that $0 \le \mathrm{H}_\epsilon(t) \le 1$ for any t.

Note that both (2.12) and (2.13) are continuous and monotonous. However, (2.13) is continuously differentiable, while (2.12) is not. Thus, (2.13) adds extra smoothness to the solution of (2.11) in comparison with (2.12) at no cost. Therefore, in all numerical examples presented in Section 4 we use (2.13), unless it is mentioned differently.

Remark 2.5. Contrary to the solution of the original problem (2.1) which does not depend on $\Psi(t),\ t < 0$, the solution of (2.11) with the Heaviside function regularized on $[-\epsilon,\epsilon]$ does depend on $\Psi(t),\ -\epsilon \le t \le 0$. Therefore, when the initial function $\Psi(t),\ t < 0$ is not known such regularization cannot be practically applied. In such case one has to use a regularization of the Heaviside function (2.2) on the time interval $[0,\epsilon]$. For instance, the analogy of (2.13) is given by:

$$\mathrm{H}_\epsilon(t) = \begin{cases} 0, & t \le 0, \\ -\dfrac{2t^3}{\epsilon^3} + \dfrac{3t^2}{\epsilon^2}, & 0 < t < \epsilon, \\ 1, & t \ge \epsilon. \end{cases} \tag{2.14}$$

We denote the vector of weighted discrepancies between the model outputs (2.11) and the measured values by $\mathrm{Z}(\theta)$. Similar to (2.3), we define the function to minimize:

$$R(\theta) = \sum_{i=1}^{N} w_i^2 (z_{c_i}(t_i, \theta) - \tilde{y}_i)^2 = \mathrm{Z}^T(\theta)\mathrm{Z}(\theta). \tag{2.15}$$

The value of θ that minimizes (2.15) is denoted by $\breve{\theta}$. Similar to (2.3), (2.15) is continuous due to the continuity of the solution of (2.11).

Differentiation of (2.15) with respect to θ_j gives

$$\frac{\partial R}{\partial \theta_j}(\theta) = 2 \sum_{i=1}^{N} (z_{c_i}(t_i, \theta) - \tilde{y}_i) \frac{\partial z_{c_i}}{\partial \theta_j}(t_i, \theta) = 2\ \mathrm{Z}^T(\theta) \frac{\partial \mathrm{Z}}{\partial \theta_j}(\theta). \tag{2.16}$$

The derivatives $\frac{\partial \mathbf{z}}{\partial \theta_j}$ can be found by solving the corresponding variational equations. The analogy of (2.6) for $\frac{\partial \mathbf{z}}{\partial \lambda_i}$ $(i = 1, 2, \ldots, m-1)$ reads

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\partial \mathbf{z}}{\partial \lambda_i}(t,\theta) &= \mathrm{H}_\epsilon(t-\tau)\left[\mathbf{f}_{\mathbf{y}}\ \dfrac{\partial \mathbf{z}}{\partial \lambda_i}(t-\tau,\theta) + \mathbf{f}_{\lambda_i}\right] \\ & \quad -\dfrac{\partial \mathrm{A}}{\partial \lambda_i}(\lambda)\ \mathbf{z}(t,\theta) - \mathrm{A}(\lambda)\ \dfrac{\partial \mathbf{z}}{\partial \lambda_i}(t,\theta), \quad 0 < t \le T, \\ \dfrac{\partial \mathbf{z}}{\partial \lambda_i}(t,\theta) &= 0, \qquad\qquad\qquad\qquad\qquad\qquad t \le 0, \end{cases} \tag{2.17}$$

which has a continuous solution $\frac{\partial \mathbf{z}}{\partial \lambda}$. So, similar to the original case, the gradient $\frac{\partial R}{\partial \lambda_i}$ is continuous in the parameters $\theta_j,\ j = 1, 2, \ldots, m$. Differentiating (2.11) with respect to τ, we obtain

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\partial \mathbf{z}}{\partial \tau}(t,\theta) &= -\delta_\epsilon(t-\tau)\ \mathbf{f} - \mathrm{A}(\lambda)\ \dfrac{\partial \mathbf{z}}{\partial \tau}(t,\theta) \\ & \quad +\mathrm{H}_\epsilon(t-\tau)\ \mathbf{f}_{\mathbf{y}}\ \left[\dfrac{\partial \mathbf{z}}{\partial \tau}(t-\tau,\theta) - \dfrac{\mathrm{d}\mathbf{z}}{\mathrm{d}t}(t-\tau,\theta)\right], \quad 0 < t \le T, \\ \dfrac{\partial \mathbf{z}}{\partial \tau}(t,\theta) &= 0, \qquad\qquad\qquad\qquad\qquad\qquad t \le 0, \end{cases} \tag{2.18}$$

where δ_ϵ is a regularization of the Dirac delta function. For instance, with regularization (2.13) of the Heaviside function the corresponding regularization of the Dirac delta function is given by

$$\delta_\epsilon(t) = \mathrm{H}'_\epsilon(t) = \begin{cases} -\dfrac{3t^2}{4\epsilon^3} + \dfrac{3}{4\epsilon}, & |t| < \epsilon, \\ 0, & |t| \ge \epsilon. \end{cases} \tag{2.19}$$

Since the right-hand side of (2.11) is continuously differentiable with respect to τ, (2.18) is defined for any $0 < t \le T$ contrary to (2.7). The solution of (2.18) is continuous. Therefore, for the regularized problem the gradient $\frac{\partial R}{\partial \tau}$ is also continuous contrary to the original case.

Similar to (2.8)-(2.9), dependent and independent confidence intervals for each estimate $\breve{\theta}_i$ $(i = 1, 2, \ldots, m)$ are given by:

$$\left[\breve{\theta}_i - \triangle_i^D,\ \breve{\theta}_i + \triangle_i^D\right], \qquad \triangle_i^D = \sqrt{\frac{m}{N-m} \frac{R(\breve{\theta})\ F_\alpha(m, N-m)}{\left(\breve{J}^T(\breve{\theta})\ \breve{J}(\breve{\theta})\right)_{ii}}} \tag{2.20}$$

and

$$\left[\breve{\theta}_i - \triangle_i^I,\ \breve{\theta}_i + \triangle_i^I\right], \qquad \triangle_i^I = \sqrt{\frac{m}{N-m}\ R(\breve{\theta})\ F_\alpha(m, N-m)\left(\left(\breve{J}^T(\breve{\theta})\ J(\breve{\theta})\right)^{-1}\right)_{ii}}, \tag{2.21}$$

respectively, where $\breve{J} = \frac{\partial \mathbb{Z}}{\partial \theta}$. The correlation coefficient between $\breve{\theta}_i$ and $\breve{\theta}_j$ is given by:

$$\breve{\rho}_{ij} = \frac{\breve{B}_{ij}}{\sqrt{\breve{B}_{ii}\breve{B}_{jj}}}, \tag{2.22}$$

where $\breve{B}(\breve{\theta}) = \left(\breve{J}^T(\breve{\theta})\ \breve{J}(\breve{\theta})\right)^{-1}$. Note that in the regularized case, (2.20)-(2.22) are unambiguously defined for any value of $\breve{\theta}$.

To conclude, model (2.1) has a discontinuous right-hand side. This discontinuity may enter into the gradient of the solution. By regularizing the right-hand side of (2.1), we achieve continuity of the gradient. It allows, firstly, to use gradient-based methods for estimation of unknown parameters and, secondly, to define statistical quantities for parameter estimates in a mathematically correct way.

3 Convergence analysis for the regularized problem

In this section we shall prove the convergence of the solution of the regularized problem (2.11) to the solution of the original problem (2.1). Assume that matrix A in the right-hand side of (2.1) is symmetric and row-diagonally dominant with positive diagonal entries [3)]. Then, it is positive definite and thus its eigenvalues are all real and positive.

Remark 3.1. For any symmetric and positive definite matrix A the estimate:

$$\|e^{-\mathrm{A}t}\| \le e^{-\mu t} \tag{3.1}$$

holds for any $t > 0$, where μ is the smallest eigenvalue of A and $\|\cdot\|$ is the L_2 norm.

The solution of (2.1), $\mathbf{y}(t,\theta)$, and the solution of (2.11), $\mathbf{z}(t,\theta)$, are both continuous for any $t \ge 0$ and θ. Therefore,

$$\mathbf{w}(t,\theta) = \mathbf{z}(t,\theta) - \mathbf{y}(t,\theta), \quad 0 \le t \le T \tag{3.2}$$

is also continuous.

Theorem 3.2. Assume that the function $\mathbf{f}$ satisfies

$$\|\mathbf{f}(t,\mathbf{u},\lambda)\| \le C, \quad \tau - \epsilon \le t \le \tau + \epsilon, \tag{3.3}$$

$$\|\mathbf{f}(t,\mathbf{u},\lambda) - \mathbf{f}(t,\mathbf{v},\lambda)\| \le L\ \|\mathbf{u}-\mathbf{v}\|, \quad 0 < t \le T \tag{3.4}$$

[3)] These assumptions are valid for the type of the applications we are interested in. Typically, matrix A includes discretized diffusion and decay terms.

for any $\mathbf{u}$ and $\mathbf{v}$. Here $C \equiv const \geq 0$, $L \equiv const > 0$, and $L \leq \mu$, with μ being the smallest eigenvalue of A. Then

$$\max_{|t-\tau|<\epsilon} \|\mathbf{w}(t,\theta)\| = O(\epsilon), \qquad \max_{|t-\tau|\geq\epsilon} \|\mathbf{w}(t,\theta)\| \leq O(\epsilon^2)$$

hold for any θ.

Proof. For $0 < t \leq \tau - \epsilon$ we have

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta) = -\mathrm{A}\mathbf{w}(t,\theta), & 0 < t \leq \tau - \epsilon, \\ \mathbf{w}(t,\theta) = 0, & t = 0, \end{cases}$$

which has the solution $\mathbf{w}(t,\theta) = 0$ for $0 \leq t \leq \tau - \epsilon$. Furthermore,

$$\mathbf{y}(t,\theta) = \mathbf{z}(t,\theta) = e^{-\mathrm{A}t}\ \Psi(0), \quad 0 \leq t \leq \tau - \epsilon.$$

For $\tau - \epsilon < t < \tau$ we have

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta) = \mathrm{H}_\epsilon(t-\tau)\ \mathbf{f}\left(t, \Psi(t-\tau), \lambda\right) - \mathrm{A}\mathbf{w}(t,\theta), & \tau - \epsilon < t < \tau, \\ \mathbf{w}(t,\theta) = 0, & t = \tau - \epsilon, \end{cases}$$

which has the solution

$$\mathbf{w}(t,\theta) = \int_{\tau-\epsilon}^{t} e^{-\mathrm{A}(t-\xi)}\ \mathrm{H}_\epsilon(\xi-\tau)\ \mathbf{f}\left(\xi, \Psi(\xi-\tau), \lambda\right) d\xi, \quad \tau - \epsilon \leq t < \tau.$$

Using the trapezoidal rule for the integral and $\mathrm{H}_\epsilon(-\epsilon) = 0$, we have

$$\mathbf{w}(t,\theta) = \frac{t-\tau+\epsilon}{2}\ \mathrm{H}_\epsilon(t-\tau)\ \mathbf{f}(t, \Psi(t-\tau), \lambda) + O((t-\tau+\epsilon)^3), \quad \tau - \epsilon \leq t < \tau.$$

Therefore,

$$\max_{\tau-\epsilon<t<\tau} \|\mathbf{w}(t,\theta)\| = O(\epsilon).$$

Using $\mathrm{H}_\epsilon(0) = \frac{1}{2} + O(\epsilon^2)$ and the continuity of $\mathbf{w}$, we have $\mathbf{w}(\tau,\theta) = \dfrac{\epsilon}{4}\ \mathbf{f}\left(\tau, \Psi(0), \lambda\right) + \Delta_1$, where $\Delta_1 = O(\epsilon^3)$.

For $\tau < t < \tau + \epsilon$ we have

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta) = \left(\mathrm{H}_\epsilon(t-\tau) - 1\right)\ \mathbf{f}\left(t, e^{-\mathrm{A}(t-\tau)}\Psi(0), \lambda\right) - \mathrm{A}\mathbf{w}(t,\theta), & \tau < t < \tau + \epsilon, \\ \mathbf{w}(t,\theta) = \dfrac{\epsilon}{4}\ \mathbf{f}\left(\tau, \Psi(0), \lambda\right) + \Delta_1, & t = \tau, \end{cases}$$

which has the solution

$$\mathbf{w}(t,\theta) = e^{-\mathrm{A}(t-\tau)}\left(\frac{\epsilon}{4}\ \mathbf{f}\left(\tau, \Psi(0), \lambda\right) + \Delta_1\right)$$

$$+ \int_{\tau}^{t} e^{-\mathrm{A}(t-\xi)}\ \left(\mathrm{H}_\epsilon(\xi-\tau) - 1\right)\ \mathbf{f}\left(\xi, e^{-\mathrm{A}(\xi-\tau)}\Psi(0), \lambda\right) d\xi, \quad \tau \leq t < \tau + \epsilon.$$

Using the trapezoidal rule for the integral and $\mathrm{H}_\epsilon(0) = \frac{1}{2} + O(\epsilon^2)$, we have

$$\mathbf{w}(t,\theta) = e^{-\mathrm{A}(t-\tau)} \left(\frac{\tau+\epsilon-t}{4} \mathbf{f}\left(\tau, \Psi(0), \lambda\right) + \Delta_1 \right)$$
$$+\frac{t-\tau}{2} \left(\mathrm{H}_\epsilon(t-\tau) - 1\right) \mathbf{f}\left(t, e^{-\mathrm{A}(t-\tau)}\Psi(0), \lambda\right) + O((t-\tau)^3), \quad \tau \le t < \tau+\epsilon.$$

So,

$$\max_{\tau \le t < \tau+\epsilon} \|\mathbf{w}(t,\theta)\| = O(\epsilon).$$

Using $\mathrm{H}_\epsilon(\epsilon) = 1$ and the continuity of $\mathbf{w}$, we have $\mathbf{w}(\tau+\epsilon,\theta) = \Delta_2$, where $\Delta_2 = e^{-\mathrm{A}\epsilon}\Delta_1 + O(\epsilon^3) = O(\epsilon^3)$.

For $\tau+\epsilon < t \le 2\tau-\epsilon$ we have

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta) = -\mathrm{A}\mathbf{w}(t,\theta), & \tau+\epsilon < t \le 2\tau-\epsilon, \\ \mathbf{w}(t,\theta) = \Delta_2, & t = \tau+\epsilon, \end{cases}$$

which has the solution

$$\mathbf{w}(t,\theta) = e^{-\mathrm{A}(t-\tau-\epsilon)}\ \Delta_2, \quad \tau+\epsilon \le t \le 2\tau-\epsilon.$$

Thus,

$$\max_{\tau+\epsilon \le t \le 2\tau-\epsilon} \|\mathbf{w}(t,\theta)\| = O(\epsilon^3),$$

and $\mathbf{w}(2\tau-\epsilon,\theta) = e^{-\mathrm{A}(\tau-2\epsilon)}\ \Delta_2$.

For $2\tau-\epsilon < t < 2\tau+\epsilon$ we have

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta) = \mathbf{f}\left(t, \mathbf{z}(t-\tau,\theta), \lambda\right) - \mathbf{f}\left(t, \mathbf{y}(t-\tau,\theta), \lambda\right) - \mathrm{A}\mathbf{w}(t,\theta), & 2\tau-\epsilon < t < 2\tau+\epsilon, \\ \mathbf{w}(t,\theta) = e^{-\mathrm{A}(\tau-2\epsilon)}\ \Delta_2, & t = 2\tau-\epsilon, \end{cases}$$

giving the solution

$$\mathbf{w}(t,\theta) = e^{-\mathrm{A}(t-\tau-\epsilon)}\ \Delta_2 + \int_{2\tau-\epsilon}^{t} e^{-\mathrm{A}(t-\xi)} \left[\mathbf{f}\left(\xi, \mathbf{z}(\xi-\tau,\theta), \lambda\right) - \mathbf{f}\left(\xi, \mathbf{y}(\xi-\tau,\theta), \lambda\right)\right] d\xi.$$

Using (3.1) and (3.4), we have

$$\|\mathbf{w}(t,\theta)\| \le e^{-\mu(t-\tau-\epsilon)}\ \|\Delta_2\| + L \int_{2\tau-\epsilon}^{t} e^{-\mu(t-\xi)} \|\mathbf{w}(\xi-\tau,\theta)\| d\xi$$
$$\le \|\Delta_2\| + 2L\epsilon \max_{\tau-\epsilon<\xi<\tau+\epsilon} \|\mathbf{w}(\xi,\theta)\| = O(\epsilon^2), \quad 2\tau-\epsilon < t < 2\tau+\epsilon.$$

Thus,

$$\max_{2\tau-\epsilon<t<2\tau+\epsilon} \|\mathbf{w}(t,\theta)\| \le O(\epsilon^2),$$

and $\mathbf{w}(2\tau+\epsilon,\theta) = \delta_1$, where $\|\delta_1\| \le O(\epsilon^2)$.

In the similar way, for $2\tau+\epsilon<t\le 3\tau$ we have

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta)=\mathbf{f}\left(t,\mathbf{z}(t-\tau,\theta),\lambda\right)-\mathbf{f}\left(t,\mathbf{y}(t-\tau,\theta),\lambda\right)-\mathrm{A}\mathbf{w}(t,\theta), & 2\tau+\epsilon<t\le 3\tau,\\ \mathbf{w}(t,\theta)=\delta_1, & t=2\tau+\epsilon, \end{cases}$$

giving the solution

$$\mathbf{w}(t,\theta)=e^{-\mathrm{A}(t-2\tau+\epsilon)}\ \delta_1+\int\limits_{2\tau+\epsilon}^{t} e^{-\mathrm{A}(t-\xi)}\left[\mathbf{f}\left(\xi,\mathbf{z}(\xi-\tau,\theta),\lambda\right)-\mathbf{f}\left(\xi,\mathbf{y}(\xi-\tau,\theta),\lambda\right)\right]d\xi.$$

Using (3.1) and (3.4), we have

$$\begin{aligned}\|\mathbf{w}(t,\theta)\|&\le e^{-\mu(t-2\tau+\epsilon)}\ \|\delta_1\|+L\int\limits_{2\tau+\epsilon}^{t} e^{-\mu(t-\xi)}\|\mathbf{w}(\xi-\tau,\theta)\|d\xi\\ &\le\|\delta_1\|+L(\tau-\epsilon)\max_{\tau+\epsilon<\xi\le 2\tau}\|\mathbf{w}(\xi,\theta)\|=O(\epsilon^2),\quad 2\tau+\epsilon<t\le 3\tau.\end{aligned}$$

Thus,

$$\max_{2\tau+\epsilon<t\le 3\tau}\|\mathbf{w}(t,\theta)\|\le O(\epsilon^2).$$

The rest we prove by induction. We assume that

$$\max_{(k-1)\tau\le t\le k\tau}\|\mathbf{w}(t,\theta)\|\le O(\epsilon^2)$$

holds for some $k\ge 3$. We have already proven it for $k=3$. Then for $k\tau<t\le(k+1)\tau$ we get

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{w}}{\mathrm{d}t}(t,\theta)=\mathbf{f}\left(t,\mathbf{z}(t-\tau,\theta),\lambda\right)-\mathbf{f}\left(t,\mathbf{y}(t-\tau,\theta),\lambda\right)-\mathrm{A}\mathbf{w}(t,\theta), & k\tau<t\le(k+1)\tau,\\ \mathbf{w}(t,\theta)=\delta_k, & t=k\tau, \end{cases}$$

where $\|\delta_k\|\le\max\limits_{(k-1)\tau\le\xi\le k\tau}\|\mathbf{w}(\xi,\theta)\|\le O(\epsilon^2)$ and we find the solution

$$\mathbf{w}(t,\theta)=e^{-\mathrm{A}(t-k\tau)}\ \delta_k+\int\limits_{k\tau}^{t} e^{-\mathrm{A}(t-\xi)}\left[\mathbf{f}\left(\xi,\mathbf{z}(\xi-\tau,\theta),\lambda\right)-\mathbf{f}\left(\xi,\mathbf{y}(\xi-\tau,\theta),\lambda\right)\right]d\xi.$$

Using (3.1), (3.4), and $L\le\mu$, we have

$$\begin{aligned}\|\mathbf{w}(t,\theta)\|&\le e^{-\mu(t-k\tau)}\ \|\delta_k\|+L\int\limits_{k\tau}^{t} e^{-\mu(t-\xi)}\|\mathbf{w}(\xi-\tau,\theta)\|d\xi\\ &\le\max_{(k-1)\tau\le\xi\le k\tau}\|\mathbf{w}(\xi,\theta)\|\left(e^{-\mu(t-k\tau)}+L\int\limits_{k\tau}^{t}e^{-\mu(t-\xi)}d\xi\right)\\ &=\max_{(k-1)\tau\le\xi\le k\tau}\|\mathbf{w}(\xi,\theta)\|\left(e^{-\mu(t-k\tau)}+\frac{L}{\mu}\left(1-e^{-\mu(t-k\tau)}\right)\right)\\ &\le\max_{(k-1)\tau\le\xi\le k\tau}\|\mathbf{w}(\xi,\theta)\|\le O(\epsilon^2),\quad k\tau\le t\le(k+1)\tau.\end{aligned}$$

Thus,

$$\max_{k\tau\le t\le(k+1)\tau}\|\mathbf{w}(t,\theta)\|\le O(\epsilon^2).$$

□

Corollary 3.3. From (2.3) and (2.15), we get

$$R(\theta)-S(\theta)=\sum_{i=1}^{N}\left(z(t_i,\theta)-y(t_i,\theta)\right)\left(z(t_i,\theta)+y(t_i,\theta)-2\tilde{y}_i\right).$$

Then, results of Theorem 3.2 imply that $R(\theta)\to S(\theta)$ as $\epsilon\to 0$ for any θ.

Corollary 3.4. $\breve{\theta}\to\hat{\theta}$ as $\epsilon\to 0$.

The obtained result in Theorem 3.2 holds only when the Heaviside function (2.2) is regularized on the time interval $[-\epsilon,\epsilon]$. If one uses the regularization of the Heaviside function on $[0,\epsilon]$, such as (2.14), then the convergence result can be derived in the following theorem[4)].

Theorem 3.5. If the Heaviside function (2.2) is regularized on the time interval $[0,\epsilon]$ and the assumptions of Theorem 3.2 hold, then

$$\max_{0<t\le T}\|\mathbf{w}(t,\theta)\|\le O(\epsilon)$$

for any θ.

4 Numerical illustrations

In this section we shall validate our findings by numerical illustrations for simple problems. In order to solve DDE model (2.1), its regularized analogy (2.11), and the corresponding variational equations (2.17)-(2.18), we use the numerical solver RADAR5 [10] based on an implicit Runge-Kutta method. We refer the reader to [11] for an extensive description of this numerical method.

For estimation of parameters in the regularized model we use the gradient-based Levenberg-Marquardt (LM) method [12]. For implementational aspects of LM method we refer the reader to [6]. Here, we only notice that the LM method is a local search approach and therefore it can get trapped in a local minimum. In order to find a global minimum we use the LM search 1000 times with initial parameter values randomly chosen from the whole search space (coarse Monte-Carlo sampling). From the obtained minima we select the one with the lowest cost function value. Since the problems we consider here for numerical illustrations have only a few parameters and a few model equations, such an approach is computationally cheap and easy to implement. Obviously, for the large-scale problems it is far more efficient to use initially a global search method in order to provide the LM search with good starting points.

[4)] The proof of this theorem follows the same lines as the proof of Theorem 3.2 and therefore, it is omitted here.

Example 4.1. We consider a simple DDE of the form:

$$\begin{cases} \dfrac{\mathrm{d}y}{\mathrm{d}t}(t) = \mathrm{H}(t-\tau)y(t-\tau) - \lambda y(t), & 0 < t \le 80, \\ y(t) = 1, & t \le 0. \end{cases} \tag{4.1}$$

The analytical solution of (4.1) is given by

$$y(t,\tau,\lambda) = \sum_{j=0}^{k} \frac{(t-j\tau)^j}{j!} e^{-\lambda(t-j\tau)}, \quad k\tau \le t < (k+1)\tau, \quad k = 0, 1, \ldots. \tag{4.2}$$

The plots on the first row of Figure 4.1 show (4.2) as a function of time for $\tau = 5$ and $\lambda = 0.5$ (left), $\lambda = 1$ (middle), and $\lambda = 1.5$ (right). As the time increases, the solution (4.2) vanishes when $\lambda > 1$, goes to non-zero steady state when $\lambda = 1$, and exponentionally grows in the case $\lambda < 1$.

We compute numerically the solution of the regularized problem, $z(t,\tau,\lambda)$, with $\epsilon = 0.01$. The plots on the second row of Figure 4.1 show $|z(t,\tau,\lambda) - y(t,\tau,\lambda)|$ as a function of time for the same parameter sets with the regularization of the Heaviside function by the cubic polynomial (2.13) on the interval $[-\epsilon, \epsilon]$. The plots confirm the results of Theorem 3.2. Note that for this problem we have $L = 1$. For $\lambda \ge L$ the absolute difference between the solution of the original problem (4.1) and its regularized analogy as predicted is of order $O(\epsilon)$ in the ϵ-neighborhood of $t = \tau$ and at most of order $O(\epsilon^2)$ elsewhere. For $\lambda < L$ the absolute difference grows in time which can be expected due to exponential growth of solution itself. However, the relative difference $|z(t,\tau,\lambda) - y(t,\tau,\lambda)|/y(t,\tau,\lambda)$ is bounded and has the same order as the absolute difference (not shown here).

The plots on the third row of Figure 4.1 show $|z(t,\tau,\lambda) - y(t,\tau,\lambda)|$ as a function of time when the Heaviside function is regularized by the cubic polynomial (2.14) on the interval $[0, \epsilon]$. The plots confirm the results of Theorem 3.5. For $\lambda \ge L$ the absolute difference is at most of order $O(\epsilon)$ everywhere.

Differentiating (4.2) with respect to λ and τ gives the analytical expression for the gradients of (4.1):

$$\frac{\partial y}{\partial \lambda}(t,\tau,\lambda) = -\sum_{j=0}^{k} \frac{(t-j\tau)^{j+1}}{j!} e^{-\lambda(t-j\tau)}, \quad k\tau \le t < (k+1)\tau, \quad k = 0, 1, \ldots \tag{4.3}$$

$$\frac{\partial y}{\partial \tau}(t,\tau,\lambda) = \begin{cases} 0, & 0 \le t < \tau, \\ \displaystyle\sum_{j=1}^{k} \frac{(t-j\tau)^{j-1}}{(j-1)!} \left(-j + \lambda(t-j\tau)\right) e^{-\lambda(t-j\tau)}, & k\tau \le t < (k+1)\tau. \end{cases} \tag{4.4}$$

We compute the gradients of the regularized problem numerically by solving the corresponding variational equations with $\epsilon = 0.01$. Figure 4.2 shows $|\frac{\partial z}{\partial \lambda} - \frac{\partial y}{\partial \lambda}|$ (first row) and $|\frac{\partial z}{\partial \tau} - \frac{\partial y}{\partial \tau}|$ (second row) as a function of time for $\tau = 5.0$ and $\lambda = 0.5$ (left), $\lambda = 1.0$ (middle), and $\lambda = 1.5$ (right). The plots reveal that for $\lambda < L$ the absolute difference between the gradients grows in time, similar to the difference between the solutions. Additionally, we observe that when $\lambda \ge L$ the difference between the gradients in λ direction is at most of

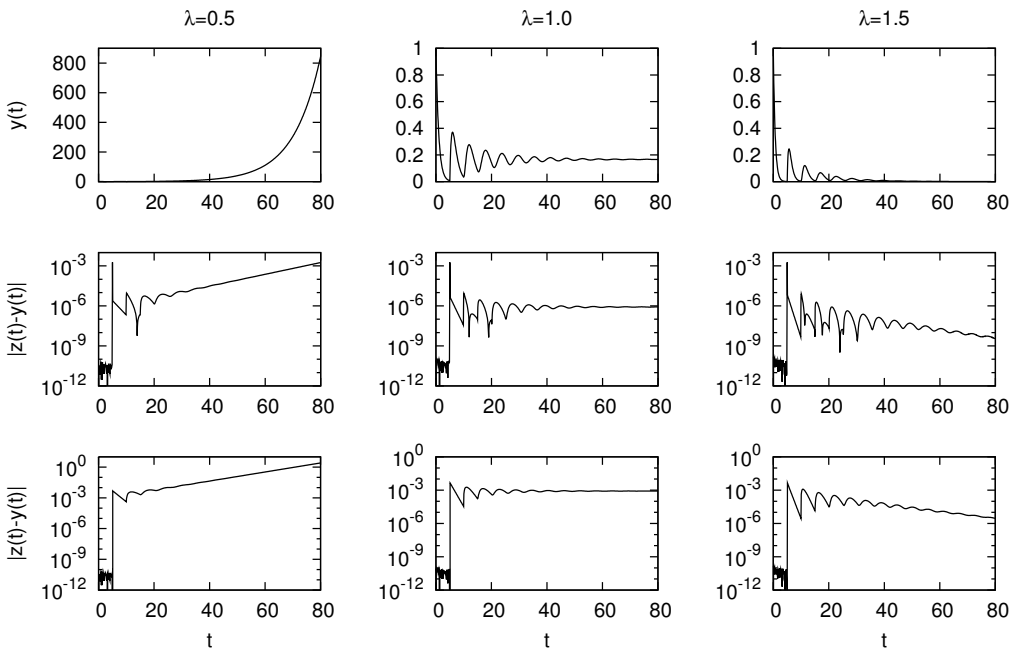

Figure 4.1: Solution (4.2) of the original problem (4.1) (first row) and the absolute difference between (4.2) and the numerical solution of the regularized problems when (2.13) (second row) and (2.14) (third row) with $\epsilon = 0.01$ are used for approximation of Heaviside function. Columns illustrate different behavior of the solution (4.2) and the regularization errors for three different values of λ; $\tau = 5$ is used in all cases.

order $O(\epsilon^2)$, while the difference between the gradients in τ direction is of order $O(1)$ in the ϵ-neighborhood of $t = \tau$, of order $O(\epsilon)$ in the ϵ-neighborhood of $t = 2\tau$, and at most of order $O(\epsilon^2)$ elsewhere. Note that the significant difference between the gradients in τ direction around $t = \tau$ is due to the discontinuity in (4.4) at that point. However, this discontinuity does not propagate in time. As the time increases, the gradient (4.4) smooths out and the order $O(\epsilon^2)$ for the difference between gradients is regained. Similar results both for $|\frac{\partial z}{\partial \lambda} - \frac{\partial y}{\partial \lambda}|$ and $|\frac{\partial z}{\partial \tau} - \frac{\partial y}{\partial \tau}|$ have been obtained with different realizations of ϵ (not shown here). The results suggest that when $\lambda \geq L$ the gradients of the regularized problem pointwisely converge to the corresponding gradients of the original problem as $\epsilon \to 0$, with the only exception for the gradient in τ direction at $t = \tau$. Moreover, they also suggest that the analogy of Theorem 3.2 holds for the difference between gradients.

Parameter Estimation We have generated an artificial dataset using solution (4.2) of problem (4.1) with $\tau^* = 5$ and $\lambda^* = 1$ at 400 time points $t = \{0.2, 0.4, \ldots, 80.0\}$. With this dataset, $\{\hat{\tau}, \hat{\lambda}\} = \{\tau^*, \lambda^*\}$ is the minimizer of the cost function (2.3) for problem (4.1) and $S(\hat{\tau}, \hat{\lambda}) = 0$. Therefore, the confidence intervals (2.8) and (2.9) for $\hat{\tau}$ and $\hat{\lambda}$ have zero size. Note that the correlation coefficient (2.10) between $\hat{\tau}$ and $\hat{\lambda}$ is not defined because the point $t = \hat{\tau}$ where (4.4) is discontinuous coincides with one of the data points and therefore the

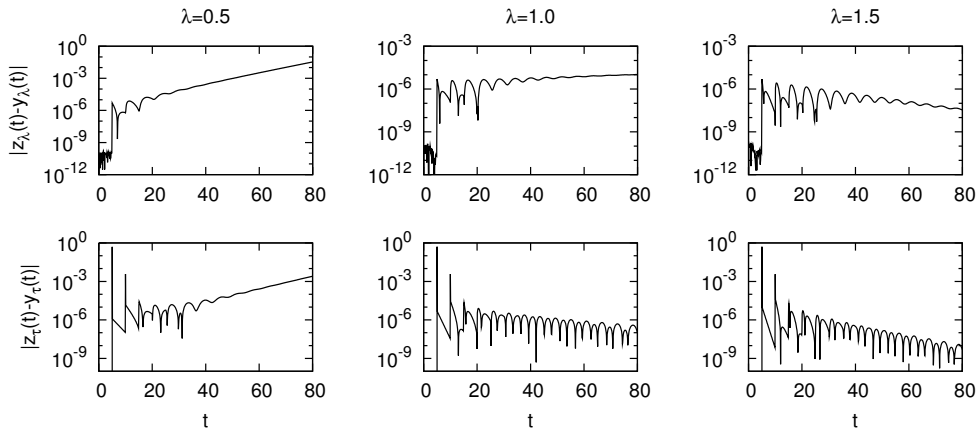

Figure 4.2: Absolute differences between the gradients of the original problem (4.1) and the regularized problem (with $\epsilon = 0.01$), $|\frac{\partial z}{\partial \lambda} - \frac{\partial y}{\partial \lambda}|$ (first row) and $|\frac{\partial z}{\partial \tau} - \frac{\partial y}{\partial \tau}|$ (second row), as a function of time with $\tau = 5.0$ and three different values of λ.

discontinuity enters into the sensitivity matrix $J(\hat{\tau}, \hat{\lambda})$.

Using this dataset, parameter estimates $\breve{\tau}$ and $\breve{\lambda}$ minimizing the cost function (2.15) for the regularized problem are obtained with the LM method. Table 4.1 nicely illustrates the first-order convergence of these parameter estimates as well as the sizes of their independent confidence intervals as $\epsilon \to 0$. The last column shows the correlation coefficient between parameter estimates computed by (2.22) for different realizations of ϵ.

ϵ	$\lvert\hat{\tau} - \breve{\tau}\rvert$	$\lvert\hat{\lambda} - \breve{\lambda}\rvert$	$\lvert\triangle^I_{\hat{\tau}} - \triangle^I_{\breve{\tau}}\rvert$	$\lvert\triangle^I_{\hat{\lambda}} - \triangle^I_{\breve{\lambda}}\rvert$	$\breve{\rho}_{\tau\lambda}$
1.0e-01	1.98e-03	2.18e-05	1.21e-03	9.96e-05	-0.16698
1.0e-02	1.78e-04	1.89e-06	1.23e-04	1.02e-05	-0.16682
1.0e-03	1.76e-05	2.31e-07	1.23e-05	1.03e-06	-0.16681
1.0e-04	1.76e-06	2.35e-08	1.24e-06	1.03e-07	-0.16680
1.0e-05	1.65e-07	2.21e-09	1.24e-07	1.03e-08	-0.16679

Table 4.1: Convergence results of parameter estimates and their statistical quantities for the regularized model when an extensive and error-free dataset is used.

Next, we have made a different dataset using solution (4.2) with $\tau^* = 5$ and $\lambda^* = 1$ at 40 time points $t = \{0.5, 1.0, \ldots, 20.0\}$ and adding to it a noise generated from a normal distribution with zero mean and standard deviation $\sigma = 0.01$. The left plot of Figure 4.3 shows the obtained data. With this dataset, $\{\hat{\tau}, \hat{\lambda}\} = \{5.001963, 1.001845\}$ is the minimizer of the cost function (2.3). It has been obtained by using the LM method for the regularized problem with $\epsilon = 10^{-9}$. Note that the LM search for the original problem (4.1) using exact expressions for the solution (4.2) and the gradients (4.3)-(4.4) has resulted in the same minimum. The independent confidence intervals for $\hat{\tau}$ and $\hat{\lambda}$ have sizes $\{\triangle^I_{\hat{\tau}}, \triangle^I_{\hat{\lambda}}\} = \{0.041787, 0.013938\}$, and the correlation coefficient between them is $\hat{\rho}_{\tau\lambda} = 0.171904$.

Using this error-dependent dataset, parameter estimates $\breve{\tau}$ and $\breve{\lambda}$ for the regularized

problem are obtained with the LM method. Table 4.2 presents the convergence results of these estimates and the corresponding statistical quantities. Contrary to the previous case, the convergence is not uniform, e.g., results for smaller values of ϵ show different convergence rate than the results for $\epsilon = 0.1$ and $\epsilon = 0.01$.

ϵ	$\lvert\hat{\tau} - \breve{\tau}\rvert$	$\lvert\hat{\lambda} - \breve{\lambda}\rvert$	$\lvert\triangle^I_{\hat{\tau}} - \triangle^I_{\breve{\tau}}\rvert$	$\lvert\triangle^I_{\hat{\lambda}} - \triangle^I_{\breve{\lambda}}\rvert$	$\lvert\hat{\rho}_{\tau\lambda} - \breve{\rho}_{\tau\lambda}\rvert$
1.0e-01	5.40e-03	4.07e-04	3.10e-03	1.59e-04	2.12e-02
1.0e-02	5.81e-04	2.94e-05	3.08e-03	4.65e-05	1.63e-02
1.0e-03	4.80e-08	5.30e-09	5.10e-09	2.27e-10	5.17e-08
1.0e-04	4.67e-10	5.39e-11	1.57e-11	5.25e-12	3.35e-11

Table 4.2: Convergence results of parameter estimates and their statistical quantities for the regularized model when the data contains measurement errors.

Note that the convergence results we have shown here do not imply that the accuracy of parameter estimates for the regularized model improves as $\epsilon \to 0$. Since the sizes of the independent confidence intervals for $\hat{\tau}$ and $\hat{\lambda}$ in this example are both of order $O(10^{-2})$, the numbers in Table 4.2 indicate that the parameter estimates $\breve{\tau}$ and $\breve{\lambda}$ have the same accuracy for all considered realizations of ϵ. Therefore, if one is interested only in the accuracy of parameter estimates, then for this example the regularization with $\epsilon = 0.1$ is sufficient. In general, however, parameter estimates and their accuracy for the original DDE model are not known and therefore the largest acceptable value of ϵ cannot be priori concluded.

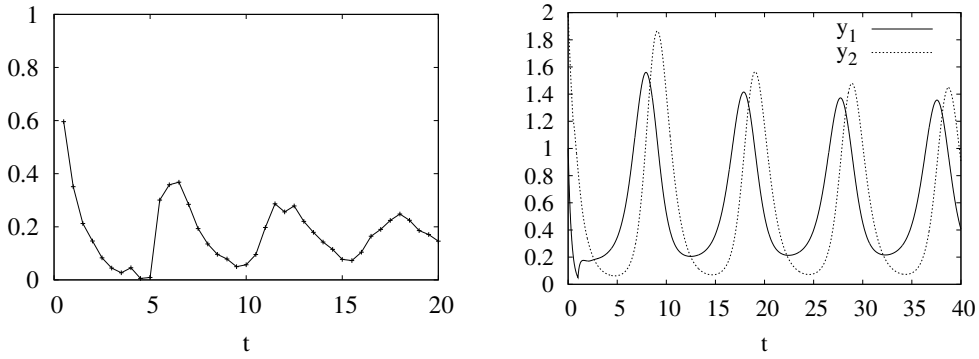

Figure 4.3: Left plot shows the data generated from solution (4.2) of problem (4.1) with $\tau^* = 5$ and $\lambda^* = 1$ by adding a normal noise to it. Right plot shows the numerical solution of problem (4.5) with $\tau_1 = 1.0$ and $\tau_2 = 0.5$.

Example 4.2. We consider DDEs of the form:

$$\begin{cases} \dfrac{\mathrm{d}y_1}{\mathrm{d}t}(t) = \mathrm{H}(t-\tau_1)\left(\dfrac{1}{2} + \dfrac{y_1^2(t-\tau_1)}{y_2(t-\tau_1)}\right) - 3y_1(t), & 0 < t \le 40, \\ \dfrac{\mathrm{d}y_2}{\mathrm{d}t}(t) = \mathrm{H}(t-\tau_2)\; y_1^2(t-\tau_2) - y_2(t), & 0 < t \le 40, \\ y_1(t) = 1, \qquad y_2(t) = 2, & t \le 0, \end{cases} \tag{4.5}$$

which originates from a Gierer-Meinhard system [13]. On the basis of this example, we shall illustrate that the convergence results hold for models having more than one time lag. The right plot of Figure 4.3 shows the numerical solution of (4.5) as a function of time for $\tau_1 = 1$ and $\tau_2 = 0.5$. The regularized analogy of (4.5) is solved numerically with three different values of ϵ. Figure 4.4 shows the absolute difference between the solutions of the original and the regularized model, $|z_1(t) - y_1(t)|$ (first row) and $|z_2(t) - y_2(t)|$ (second row). As we can see, the difference is at most of order $O(\epsilon^2)$ everywhere except in the vicinity of time points $t = \tau_1$ and $t = \tau_2$, where the corresponding components of the systems differ from each other by $O(\epsilon)$. These results support the convergence results obtained for the problem in Example 4.1 as well as the theoretical results of Theorem 3.2.

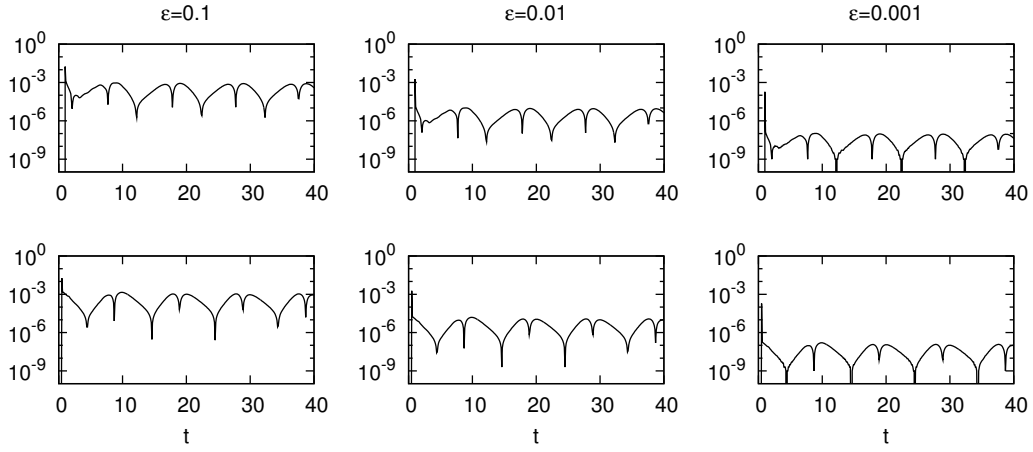

Figure 4.4: Absolute difference between the solution of problem (4.5) and the solution of the corresponding regularized problem as a function of time, i.e., $|z_1(t) - y_1(t)|$ (first row) and $|z_2(t) - y_2(t)|$ (second row). Here, $\tau_1 = 1$, $\tau_2 = 0.5$, and the regularized problem is solved with three different values of ϵ.

5 Conclusions

In this note we have studied DDE models with a right-hand side being discontinuous in time and time-lag parameters. For these type of DDEs the corresponding gradients of the solution are discontinuous as well. This causes difficulties in the parameter estimation procedures or the determinability analysis of parameter estimates when any of those methods is based on gradient information.

We have suggested to regularize the model to make it continuous. The regularization is applied at an ϵ-neighborhood of the discontinuity points. The idea of regularization is to obtain a solution which is sufficiently close to the solution of the original model almost everywhere in time and has extra smoothness in the vicinity of the discontinuity points.

For such a regularization strategy, we have shown that the solution of the regularized model converges to the solution of the original problem as $\epsilon \to 0$ for the DDEs under consideration. Moreover, we have been able to derive the convergence rate. Additionally, our

theoretical results imply that the parameter estimates obtained for the regularized model converge to the corresponding estimates of the original problem. We have been able to confirm our findings by numerical illustrations for two simple test problems.

Finally, we note that the type of DDEs considered here finds an application in developmental biology. Namely, our work was inspired by the model of spatio-temporal pattern formation for gene expression at the early developmental time of the fruit-fly *Drosophila* which incorporates the delay in the protein production. The aim of our future work is to use a regularized model for this real-life biological problem, to estimate model parameters using a large-scale dataset, and to apply statistical analysis for the obtained parameter estimates.

Acknowledgment I acknowledge support from NWO's 'Computational Life Science' program, projectnr. 635.100.010. I would like to thank Prof. dr. J.G. Verwer and J.G. Blom for their valuable comments and suggestions.

References

[1] G. A. Bocharov, F. A. Rihan (2000), *Numerical modelling in biosciences using delay differential equations*, J. Comput. Appl. Math. 125, pp. 183-199.

[2] P. W. Nelson, J. D. Murray, A. S. Perelson (2000), *A model of HIV pathogenesis that includes an intracellular delay*, Math. Biosciences 163 (2), pp. 201-215.

[3] A. Bellen, M. Zennaro (2003), *Numerical methods for delay differential equations*, Clarendon, Oxford.

[4] C. T. H. Baker, C. A. H. Paul (1997), *Pitfalls in Parameter Estimation for Delay Differential Equations*, SIAM J. Sci. Comput. 18 (1), pp. 305-314.

[5] F. H. Clarke (1983), *Optimization and Nonsmooth Analysis*, Wiley, New York.

[6] M. Ashyraliyev, J. Jaeger, J. G. Blom (2008), *On Parameter Estimation and Determinability for Drosophila Gap Gene Circuits*, BMC Systems Biology 2:83.

[7] Manu (2007), *Canalization of Gap Gene Expression During Early Development in Drosophila melanogaster*, PHD thesis, Stony Brook University.

[8] C. T. H. Baker, C. A. H. Paul (2006), *Discontinuous solutions of neutral delay differential equations*, Appl. Numer. Math. 56, pp. 284-304.

[9] G. A. F. Seber, C. J. Wild (1988), *Nonlinear regression*, New York, John Wiley & Sons.

[10] N. Guglielmi, E. Hairer (2001), *Implementing Radau IIA Methods for Stiff Delay Differential Equations*, Computing 67, pp. 1-12.

[11] N. Guglielmi, E. Hairer (2005), *Users guide for the code RADAR5 - version 2.1*, Technical report, Universita dellAquila, Italy.

[12] D. W. Marquardt (1963), *An algorithm for least-squares estimation of nonlinear parameters*, SIAM J. Appl. Math. 11, pp. 431-441.

[13] J. D. Murray (2002), *Mathematical Biology*, Berlin, Springer.

Centrum Wiskunde & Informatica (CWI) is the national research institute for mathematics and computer science in the Netherlands. The institute's strategy is to concentrate research on four broad, societally relevant themes: earth and life sciences, the data explosion, societal logistics and software as service.

Centrum Wiskunde & Informatica (CWI) is het nationale onderzoeksinstituut op het gebied van wiskunde en informatica. De strategie van het instituut concentreert zich op vier maatschappelijk relevante onderzoeksthema's: aard- en levenswetenschappen, de data-explosie, maatschappelijke logistiek en software als service.

Bezoekadres:
Science Park 123
Amsterdam

Postadres:
Postbus 94079, 1090 GB Amsterdam
Telefoon 020 592 41 81
Fax 020 592 41 99
info@cwi.nl
www.cwi.nl

CWI

Centrum Wiskunde & Informatica