

PAPER

On iteratively regularized predictor–corrector algorithm for parameter identification*

To cite this article: Alexandra Smirnova and Anatoly Bakushinsky 2020 *Inverse Problems* **36** 125015

View the [article online](#) for updates and enhancements.

You may also like

- [Automatic two-channel sleep staging using a predictor–corrector method](#)
S Riaz, T Wendler and J Pilz
- [Goal oriented adaptivity in the IRGNM for parameter identification in PDEs: II. all-at-once formulations](#)
B Kaltenbacher, A Kirchner and B Vexler
- [An optimized linearization-based predictor–corrector algorithm for the numerical simulation of nonlinear FDEs](#)
Zaid Odibat and Nabil Shawagfeh



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

On iteratively regularized predictor–corrector algorithm for parameter identification^{*}

Alexandra Smirnova^{1,**}  and Anatoly Bakushinsky²

¹ Department of Mathematics and Statistics, Georgia State University, Atlanta, United States of America

² Federal Research Center, ‘Computer Science and Control’ of Russian Academy of Sciences, Mari State University, Russia

E-mail: asmirnova@gsu.edu and bakush@isa.ru

Received 24 June 2020, revised 14 October 2020

Accepted for publication 27 October 2020

Published 8 December 2020



CrossMark

Abstract

We study a constrained optimization problem of stable parameter estimation given some noisy (and possibly incomplete) measurements of the state observation operator. In order to find a solution to this problem, we introduce a hybrid regularized predictor–corrector scheme that builds upon both, all-at-once formulation, recently developed by B. Kaltenbacher and her co-authors, and the so-called traditional route, pioneered by A. Bakushinsky. Similar to all-at-once approach, our proposed algorithm does not require solving the constraint equation numerically at every step of the iterative process. At the same time, the predictor–corrector framework of the new method avoids the difficulty of dealing with large solution spaces resulting from all-at-once make-up, which inevitably leads to oversized Jacobian and Hessian approximations. Therefore our predictor–corrector algorithm (PCA) has the potential to save time and storage, which is critical when multiple runs of the iterative scheme are carried out for uncertainty quantification. To assess numerical efficiency of novel PCA, two parameter estimation inverse problems in epidemiology are considered. All experiments are carried out with real data on COVID-19 pandemic in Netherlands and Spain.

Keywords: iterative regularization, epidemiology, stable parameter estimation

^{*} Supported by NSF awards 1818886 and 2011622 (DMS Computational Mathematics) and Russian Science Foundation project 20-11-20085.

^{**} Author to whom any correspondence should be addressed.

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. Scope

The topic of this research has been inspired by numerous challenges of stable parameter estimation in epidemiology. However, similar inverse problems arise in acoustic sensing, signal and image processing, biomedical imaging, gravitational sounding, hydrology, and other fields [1, 5, 8]. Therefore in what follows we present our proposed algorithm in a general framework of constrained least squares problems in Hilbert spaces.

Real-time reconstruction of disease parameters for an emerging outbreak helps to provide crucial information for the design of public health policies and control measures. Regardless of a particular disease, fitting model predictions for an invading pathogen to reported incidence series yields an ill-posed problem due to excessive noise propagation coupled with unavoidable delays in processing of epidemic data. In order to solve this ill-posed problem in a stable manner, regularized Gauss–Newton or Levenberg–Marquardt algorithms [3, 4, 14, 18, 22, 26, 27, 29, 32] are commonly used to minimize the cost functional. Oftentimes, the biological model (which may be a system of nonlinear ordinary or partial differential equations), constraining the function minimization problem, does not have a closed form-solution and has to be solved numerically at every step of the iterative process. This can noticeably increase the computational complexity of parameter estimation, while at the same time making it even more sensitive to the presence of noise in the input data. The goal of this paper is to construct a regularized predictor–corrector algorithm (PCA) that will mitigate excessive computational cost of a quasi-Newton step and, by doing so, incorporate extra layer of stability in the iterative process.

1.2. Constrained least squares problem

To achieve this goal, we consider a general problem of stable parameter estimation from an operator equation connecting the unknown parameter, θ , to a state variable, u ,

$$G(\theta, u) = g, \quad G: \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}, \quad (1.1)$$

given some noisy (and possibly incomplete) measurements of the state observation operator, B :

$$B(u) = d, \quad \|d - d_\delta\| \leq \delta, \quad B: \mathcal{Y} \rightarrow \mathcal{H}, \quad (1.2)$$

with $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{H}$ being some Hilbert spaces. Thus our goal is to solve the constrained least squares problem (CLSP)

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|B(u) - d\|^2 \quad \text{with respect to } u \\ & \text{subject to} \quad G(\theta, u) = g. \end{aligned} \quad (1.3)$$

An all-at-once formulation of CLSP (1.3), as recently introduced by Kaltenbacher and her co-authors in [15–17], is to solve the combined equation

$$F(q) = f, \quad q := \begin{pmatrix} \theta \\ u \end{pmatrix}, \quad F(q) := \begin{pmatrix} G(\theta, u) \\ B(u) \end{pmatrix}, \quad (1.4)$$

$$f := \begin{pmatrix} g \\ d \end{pmatrix}, \quad F: \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z} \times \mathcal{H}, \quad (1.5)$$

by some regularized computational method. This is an alternative to a more traditional approach, which consists in reducing (1.1) and (1.2) to the operator equation (or to the corresponding least squares) over the parameter space only:

$$J(\theta) = d, \quad J: \mathcal{X} \rightarrow \mathcal{H}. \quad (1.6)$$

To arrive at (1.6), one solves equation (1.1) for the parameter-to-state map, $u = u(\theta)$, and then substitutes it into (1.2). Once (1.6) has been derived, θ is computed by a regularized optimization algorithm but over a smaller solution space as compared to the solution space in all-at-once formulation. Thus in (1.6), J is the composition of the parameter-to-state map, $u = u(\theta)$, satisfying

$$G(\theta, u(\theta)) = g, \quad u: \mathcal{X} \rightarrow \mathcal{Y}, \quad (1.7)$$

and the observation operator, B , acting between $\mathcal{Y}$ and $\mathcal{H}$. That is, $J(\theta) := B(u(\theta))$. As mentioned above, it is not always possible to solve (1.1) for u analytically. Therefore, in many cases, one has to find $u_k = u(\theta_k)$ numerically for each current value, θ_k , of the unknown parameter.

1.3. Proposed algorithm

In this paper, we propose a hybrid method that builds upon both all-at-once and traditional strategies. It employs a predictor–corrector kind of algorithm, where one updates θ while freezing u , and then u is modified while θ is kept unchanged. More specifically, given $\begin{pmatrix} \theta_k \\ u_k \end{pmatrix}$, one transitions from θ_k to θ_{k+1} by applying one step of the modified iteratively regularized Gauss–Newton (MIRGN) procedure [3, 4, 18, 26, 27]:

$$\begin{aligned} \theta_{k+1} &= \theta_k - [G_\theta'^*(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} \\ &\quad \times \{G_\theta'^*(\theta_k, u_k)(G(\theta_k, u_k) - g) + \alpha_k T^*T(\theta_k - \bar{\theta})\}. \end{aligned} \quad (1.8)$$

Then, given $\begin{pmatrix} \theta_{k+1} \\ u_k \end{pmatrix}$, one computes u_{k+1} using classical Gauss–Newton scheme [19]

$$\begin{aligned} u_{k+1} &= u_k - [G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k) + B'^*(u_k)B'(u_k)]^{-1} \\ &\quad \times \{G_u'^*(\theta_{k+1}, u_k)(G(\theta_{k+1}, u_k) - g) + B'^*(u_k)(B(u_k) - d_\delta)\}. \end{aligned} \quad (1.9)$$

Note that MIRGN scheme (1.8) originates from variational regularization [29, 33] in the form

$$\min_{\theta \in \mathcal{X}} \left\{ \frac{1}{2} \|G(\theta, u_k) - g\|^2 + \frac{\alpha_k}{2} \|T(\theta - \bar{\theta})\|^2 \right\}. \quad (1.10)$$

For iteration (1.8) to be well-defined, we assume that T is surjective linear operator between two Hilbert spaces, $\mathcal{X}$ and $\mathcal{W}$, satisfying the condition [27]: for any $h \in \mathcal{X}$,

$$(T^*Th, h) \geq \lambda \|h\|^2, \quad \lambda > 0. \quad (1.11)$$

Method (1.9), on the other hand, is the classical Gauss–Newton algorithm applied to the nonlinear minimization problem

$$\min_{u \in \mathcal{Y}} \left\{ \frac{1}{2} \|G(\theta_{k+1}, u) - g\|^2 + \frac{1}{2} \|B(u) - d\|^2 \right\}, \quad \|d - d_\delta\| \leq \delta. \quad (1.12)$$

Gauss–Newton procedure (1.9) does not need to be regularized, since minimizing the functional $\|G(\theta_{k+1}, u) - g\|^2$ with respect to $u \in \mathcal{Y}$ is not, generally, an ill-posed problem: it may be a forward problem in ordinary or partial differential equations, for example.

1.4. Convergence results

In order to proceed with our convergence analysis, we make the following assumptions.

Condition A. Let the operator $G(\theta, u)$ in (1.1) be Fréchet differentiable in $\mathcal{O}_\eta(\hat{q})$ with respect to u and

$$s := \sup_{\begin{pmatrix} \theta \\ u \end{pmatrix} \in \mathcal{O}_\eta(\hat{q})} \|(G_u^*(\theta, u)G_u'(\theta, u))^{-1}\| < \infty. \quad (1.13)$$

Suppose also that for any $\begin{pmatrix} \theta \\ u_1 \end{pmatrix}$ and $\begin{pmatrix} \theta \\ u_2 \end{pmatrix} \in \mathcal{O}_\eta(\hat{q})$, the operator $G_u'(\theta, u)$ is bounded and Lipschitz-continuous in u . That is, there is $C > 0$, such that

$$\|G_u'(\theta, u_1)\| \leq C \quad \text{and} \quad \|G_u'(\theta, u_1) - G_u'(\theta, u_2)\| \leq C\|u_1 - u_2\|, \quad \begin{pmatrix} \theta \\ u_1 \end{pmatrix}, \begin{pmatrix} \theta \\ u_2 \end{pmatrix} \in \mathcal{O}_\eta(\hat{q}). \quad (1.14)$$

Here $\hat{q} := \begin{pmatrix} \hat{\theta} \\ \hat{u} \end{pmatrix} \in \mathcal{X} \times \mathcal{Y}$ is a solution to (1.1) and (1.2), which is not necessarily unique, and $\mathcal{O}_\eta(\hat{q}) := \{q \in \mathcal{X} \times \mathcal{Y} : \|q - \hat{q}\| \leq \eta\}$ with radius $\eta > 0$ specified in (2.2) below.

Naturally, for any $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{X} \times \mathcal{Y}$, we define

$$\left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|_{\mathcal{X} \times \mathcal{Y}}^2 := \|x\|_{\mathcal{X}}^2 + \|y\|_{\mathcal{Y}}^2. \quad (1.15)$$

Condition B. Let the operator $G(\theta, u)$ in (1.1) be Fréchet differentiable in $\mathcal{O}_\eta(\hat{q})$ with respect to θ and $G_\theta'(\theta, u)$ be Lipschitz-continuous in both θ and u . That is, for any $\begin{pmatrix} \theta_1 \\ u \end{pmatrix}, \begin{pmatrix} \theta_2 \\ u \end{pmatrix}, \begin{pmatrix} \theta \\ u_1 \end{pmatrix}$, and $\begin{pmatrix} \theta \\ u_2 \end{pmatrix} \in \mathcal{O}_\eta(\hat{q})$

$$\|G_\theta'(\theta_1, u) - G_\theta'(\theta_2, u)\| \leq C\|\theta_1 - \theta_2\| \quad \text{and} \quad \|G_\theta'(\theta, u_1) - G_\theta'(\theta, u_2)\| \leq C\|u_1 - u_2\|. \quad (1.16)$$

Condition C. Let the operator $B(u)$ in (1.2) be Fréchet differentiable and Lipschitz-continuous in $\mathcal{O}_\eta(\hat{q})$. That is, there is $C > 0$, such that

$$\|B_u'(u_1) - B_u'(u_2)\| \leq C\|u_1 - u_2\| \quad \text{for any} \quad \begin{pmatrix} \theta \\ u_1 \end{pmatrix}, \begin{pmatrix} \theta \\ u_2 \end{pmatrix} \in \mathcal{O}_\eta(\hat{q}). \quad (1.17)$$

Condition D. Let for the regularization sequence, $\{\alpha_k\}$, the following assumptions are met

$$\alpha_k \geq \alpha_{k+1} > 0, \quad a := \sup_{k=0,1,2,\dots} \sqrt{\frac{\alpha_k}{\alpha_{k+1}}} < \infty, \quad \lim_{k \rightarrow \infty} \alpha_k = 0. \quad (1.18)$$

In section 2, using the above four conditions, we justify local convergence of algorithm (1.8) and (1.9), first, in the noise-free case (theorem 2.5) and, second, for noise-contaminated input data (theorem 2.7). In the noise-free case, for θ_0 and u_0 sufficiently close to $\hat{\theta}$ and $\hat{u}$, respectively, it is proven in theorem 2.5 that there are positive constants, l_1 and l_2 such that

$$\|\theta_k - \hat{\theta}\| \leq l_1 \sqrt{\alpha_k} \quad \text{and} \quad \|u_k - \hat{u}\| \leq l_2 \alpha_k, \quad k = 1, 2, \dots \quad (1.19)$$

For noise-contaminated data, as shown in theorem 2.7, *a priori* stopping rule (2.46) yields the following convergence rates:

$$\|\theta_{\mathcal{K}(\delta)} - \hat{\theta}\| = O(\sqrt{\delta}) \quad \text{and} \quad \|u_{\mathcal{K}(\delta)} - \hat{u}\| = O(\delta) \quad (1.20)$$

with $\mathcal{K}(\delta) \rightarrow \infty$ as $\delta \rightarrow 0$.

1.5. Related Work and outline of the paper

In a well-posed case, the idea of hybrid algorithms goes back to Newton–Gauss–Seidel iterative process [19, 21]. In this process, the Newton step, p_k , which needs to be found from a system of linear equations for every k , is calculated approximately by Gauss–Seidel iterative scheme [21]. Of course, for nonlinear optimization, Newton’s method can be replaced with Gauss–Newton or BFGS algorithms. On the other hand, the linear system for p_k can be solved by Landweber, conjugate gradient, or some other iterative procedure. In the ill-posed case, inner iterations, apart from solving the linear system or equations, have also been used to (iteratively) regularize a Newton-type scheme [13, 18, 20].

The manner in which iterations are constructed in predictor–corrector algorithm (1.8) and (1.9) is different from the Newton–Gauss–Seidel process (NGSP). First, algorithm (1.8) and (1.9) is coupled with iterative regularization [4, 18], which makes it stable with respect to noise in the input data. Second, unlike NGSP, method (1.8) and (1.9) is not a combination of primary (nonlinear) and secondary (linear) iterations. Instead, the two alternating iterative sequences are generated in such a way that allows to simultaneously minimize the cost functional and to approximate a solution to the (non)linear constraint equation. Therefore, like in the case of all-at-once approach [15–17], the proposed method (1.8) and (1.9) does not require solving the constraint equation at every step, k , of the iterative process. For PDE constraints, encountered in various applied problems, the need to solve (1.1) for $u_k = u(\theta_k)$ can be a major bottleneck in practical implementation of the ‘traditional’ regularized procedure. On the other hand, in epidemiology, the constraining ODEs are often stiff and hard to solve in a stable manner. Thus, an ability to eliminate the computation of $u_k = u(\theta_k)$ from (1.1) may be of significant advantage in many cases. At the same time, the predictor–corrector framework of algorithm (1.8) and (1.9) avoids the difficulty of dealing with large solution spaces resulting from all-at-once make-up, which inevitably leads to oversized Jacobian and Hessian approximations. Thus our new method (1.8) and (1.9) has the potential to save time and storage, which is critical when multiple runs of the iterative scheme are carried out for uncertainty quantification.

The paper is organized as follows. In section 2, some auxiliary lemmas are formulated and proven, and the main convergence results, theorems 2.5 and 2.7, are stated and justified. In theorem 2.5, a noise-free case is studied, while in theorem 2.7 an *a priori* stopping rule for iterative process (1.8) and (1.9) with noise-contaminated data is given leading to convergence rates in both u and θ . Numerical experiments on stable parameter estimation and forecasting in epidemiology with real data on COVID-19 pandemic in Netherlands and Spain are presented in section 3. Conclusions and future plans are outlined in section 4.

2. Proofs of main results

2.1. Auxiliary lemmas

In this subsection, as a preliminary step, we carry out the convergence analysis of algorithm (1.8) and (1.9) under the assumption that the data is clean, i.e., $d = d_\delta$. We begin with the following lemma.

Lemma 2.1. *Suppose conditions A and B, and (1.11) are met, the point $\begin{pmatrix} \theta_k \\ u_k \end{pmatrix} \in \mathcal{O}_\eta(\hat{q})$, and the test function, $\bar{\theta}$, is selected in such a way that*

$$T^*T(\hat{\theta} - \bar{\theta}) = G_\theta^{f*}(\hat{\theta}, \hat{u})v, \quad v \in \mathcal{Z}, \quad \|v\| \leq \varepsilon. \quad (2.1)$$

Then for $\{\theta_k\}$ defined in (1.8), one has

$$\begin{aligned} \theta_{k+1} - \hat{\theta} = & -[G_\theta^{f*}(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} G_\theta^{f*}(\theta_k, u_k) \{ \mathcal{B}(\theta_k, \hat{\theta}, u_k, \hat{u}) + \alpha_k v \} \\ & - \alpha_k [G_\theta^{f*}(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} \{ G_\theta'(\hat{\theta}, \hat{u}) - G_\theta'(\theta_k, u_k) \} v, \end{aligned} \quad (2.2)$$

where

$$\|\mathcal{B}(\theta_k, \hat{\theta}, u_k, \hat{u})\| \leq C\|u_k - \hat{u}\| + \frac{C}{2}\|\theta_k - \hat{\theta}\|^2, \quad (2.3)$$

and the following estimate holds

$$\begin{aligned} \|\theta_{k+1} - \hat{\theta}\| \leq & \frac{1}{2\sqrt{\lambda\alpha_k}} \left[C\|u_k - \hat{u}\| + \frac{C}{2}\|\theta_k - \hat{\theta}\|^2 + \alpha_k \varepsilon \right] \\ & + \frac{C\varepsilon}{\lambda} (\|u_k - \hat{u}\| + \|\theta_k - \hat{\theta}\|). \end{aligned} \quad (2.4)$$

Proof. Taking into account condition B, one derives

$$G(\hat{\theta}, u_k) - g = G(\theta_k, u_k) - g + G_\theta'(\theta_k, u_k)(\hat{\theta} - \theta_k) + \mathcal{B}_1(\theta_k, \hat{\theta}, u_k), \quad (2.5)$$

$$\|\mathcal{B}_1(\theta_k, \hat{\theta}, u_k)\| \leq \frac{C}{2}\|\theta_k - \hat{\theta}\|^2. \quad (2.6)$$

From identity (2.5) coupled with condition A, one concludes

$$\begin{aligned} G(\theta_k, u_k) - g &= G(\hat{\theta}, u_k) - g + G_\theta'(\theta_k, u_k)(\theta_k - \hat{\theta}) - \mathcal{B}_1(\theta_k, \hat{\theta}, u_k) \\ &= G(\hat{\theta}, \hat{u}) - g + \mathcal{B}_2(\hat{\theta}, u_k, \hat{u}) + G_\theta'(\theta_k, u_k)(\theta_k - \hat{\theta}) - \mathcal{B}_1(\theta_k, \hat{\theta}, u_k), \end{aligned} \quad (2.7)$$

$$\|\mathcal{B}_2(\hat{\theta}, u_k, \hat{u})\| \leq C\|u_k - \hat{u}\|. \quad (2.8)$$

Since $G(\hat{\theta}, \hat{u}) - g = 0$, representation (2.7) implies

$$\begin{aligned} \theta_{k+1} - \hat{\theta} = & -[G_\theta^{f*}(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} G_\theta^{f*}(\theta_k, u_k) \mathcal{B}(\theta_k, \hat{\theta}, u_k, \hat{u}) \\ & - \alpha_k [G_\theta^{f*}(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} T^*T(\hat{\theta} - \bar{\theta}), \end{aligned} \quad (2.9)$$

where $\mathcal{B}(\theta_k, \hat{\theta}, u_k, \hat{u})$ is defined as

$$\mathcal{B}(\theta_k, \hat{\theta}, u_k, \hat{u}) := -\mathcal{B}_1(\theta_k, \hat{\theta}, u_k) + \mathcal{B}_2(\hat{\theta}, u_k, \hat{u}). \quad (2.10)$$

Equalities (2.9) and (2.10) together with source condition (2.1) yield (2.2) and (2.3). This completes the first part of the proof. Following the same argument as in [27], one obtains

$$[G_\theta'^*(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} = (T^*T)^{-1/2}[A_k^*A_k + \alpha_k I]^{-1}(T^*T)^{-1/2}, \quad (2.11)$$

where $A_k := G_\theta'(\theta_k, u_k)(T^*T)^{-1/2}$. Therefore by assumption (1.11),

$$\begin{aligned} & \| [G_\theta'^*(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} G_\theta'^*(\theta_k, u_k) \| \\ & \leq \| (T^*T)^{-1/2} [A_k^*A_k + \alpha_k I]^{-1} (A_k^*A_k)^{1/2} \| \\ & \leq \frac{1}{\sqrt{\lambda}} \sup_{t \in \sigma(A_k^*A_k)} \frac{\sqrt{t}}{t + \alpha} \leq \frac{1}{2\sqrt{\lambda\alpha_k}}. \end{aligned} \quad (2.12)$$

In (2.12), $\sigma(A_k^*A_k) \subseteq [0, \|A_k\|^2]$ is the spectrum of the self-adjoint operator $A_k^*A_k$. Additionally, it follows from (2.11) that

$$\alpha_k \| [G_\theta'^*(\theta_k, u_k)G_\theta'(\theta_k, u_k) + \alpha_k T^*T]^{-1} \| \leq \frac{1}{\lambda}. \quad (2.13)$$

Combining (2.1), (2.2) with (2.12) and (2.13), one gets

$$\|\theta_{k+1} - \hat{\theta}\| \leq \frac{1}{2\sqrt{\lambda\alpha_k}} \left[C\|u_k - \hat{u}\| + \frac{C}{2}\|\theta_k - \hat{\theta}\|^2 + \alpha_k \varepsilon \right] + \frac{\varepsilon}{\lambda} (\|G_\theta'(\hat{\theta}, \hat{u}) - G_\theta'(\theta_k, u_k)\|). \quad (2.14)$$

Inequality (2.14) along with the representation

$$G_\theta'(\hat{\theta}, \hat{u}) - G_\theta'(\theta_k, u_k) = [G_\theta'(\hat{\theta}, \hat{u}) - G_\theta'(\hat{\theta}, u_k)] + [G_\theta'(\hat{\theta}, u_k) - G_\theta'(\theta_k, u_k)] \quad (2.15)$$

imply that estimate (2.4) is true as claimed.

In order to derive the upper bound for $\|u_{k+1} - \hat{u}\|$, we now prove the following three lemmas.

Lemma 2.2. Suppose conditions A–C are met, and $\begin{pmatrix} \theta_k \\ u_k \end{pmatrix}$ and $\begin{pmatrix} \theta_{k+1} \\ u_k \end{pmatrix} \in \mathcal{O}_\eta(\hat{q})$. Then for $\{u_k\}$ defined in (1.9), one has

$$\begin{aligned} u_{k+1} - \hat{u} &= -[G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k) + B'^*(u_k)B'(u_k)]^{-1} \left\{ G_u'^*(\theta_{k+1}, u_k) \right. \\ & \quad \left. + [G_\theta'(\theta_k, u_k)(\theta_{k+1} - \hat{\theta}) + \mathcal{P}(\theta_k, \theta_{k+1}, \hat{\theta}, u_k, \hat{u})] + B'^*(u_k)\mathcal{L}(u_k, \hat{u}) \right\}, \end{aligned} \quad (2.16)$$

where

$$\begin{aligned} \|\mathcal{P}(\theta_k, \theta_{k+1}, \hat{\theta}, u_k, \hat{u})\| &\leq C \left\{ \frac{\|u_k - \hat{u}\|^2}{2} + \frac{\|\theta_{k+1} - \hat{\theta}\|^2}{2} \right. \\ & \quad \left. + (\|\theta_k - \hat{\theta}\| + \|u_k - \hat{u}\|) \|\theta_{k+1} - \hat{\theta}\| \right\}, \\ \|\mathcal{L}(u_k, \hat{u})\| &\leq \frac{C}{2} \|u_k - \hat{u}\|^2. \end{aligned} \quad (2.17)$$

Proof. Condition A yields

$$G(\theta_{k+1}, \hat{u}) - g = G(\theta_{k+1}, u_k) - g + G'_u(\theta_{k+1}, u_k)(\hat{u} - u_k) + \mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k), \quad (2.18)$$

$$\|\mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k)\| \leq \frac{C}{2} \|u_k - \hat{u}\|^2. \quad (2.19)$$

From expression (2.19) and condition B, one obtains

$$\begin{aligned} G(\theta_{k+1}, u_k) - g &= G(\theta_{k+1}, \hat{u}) - g + G'_u(\theta_{k+1}, u_k)(u_k - \hat{u}) - \mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k) \\ &= G(\hat{\theta}, \hat{u}) - g + G'_\theta(\hat{\theta}, \hat{u})(\theta_{k+1} - \hat{\theta}) + \mathcal{E}_2(\theta_{k+1}, \hat{\theta}, \hat{u}) \\ &\quad + G'_u(\theta_{k+1}, u_k)(u_k - \hat{u}) - \mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k), \end{aligned} \quad (2.20)$$

$$\|\mathcal{E}_2(\theta_{k+1}, \hat{\theta}, \hat{u})\| \leq \frac{C}{2} \|\theta_{k+1} - \hat{\theta}\|^2. \quad (2.21)$$

Since $\begin{pmatrix} \hat{\theta} \\ \hat{u} \end{pmatrix} \in \mathcal{X} \times \mathcal{Y}$ is a solution to (1.1) and (1.2), one gets

$$\begin{aligned} G(\theta_{k+1}, u_k) - g &= G'_u(\theta_{k+1}, u_k)(u_k - \hat{u}) + G'_\theta(\hat{\theta}, \hat{u})(\theta_{k+1} - \hat{\theta}) - \mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k) \\ &\quad + \mathcal{E}_2(\theta_{k+1}, \hat{\theta}, \hat{u}) = G'_u(\theta_{k+1}, u_k)(u_k - \hat{u}) + G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta}) \\ &\quad + (G'_\theta(\hat{\theta}, \hat{u}) - G'_\theta(\theta_k, u_k))(\theta_{k+1} - \hat{\theta}) + (G'_\theta(\hat{\theta}, u_k) \\ &\quad - G'_\theta(\theta_k, u_k))(\theta_{k+1} - \hat{\theta}) - \mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k) + \mathcal{E}_2(\theta_{k+1}, \hat{\theta}, \hat{u}). \end{aligned} \quad (2.22)$$

If one introduces the notation:

$$\begin{aligned} \mathcal{P}(\theta_k, \theta_{k+1}, \hat{\theta}, u_k, \hat{u}) &:= (G'_\theta(\hat{\theta}, \hat{u}) - G'_\theta(\hat{\theta}, u_k))(\theta_{k+1} - \hat{\theta}) + (G'_\theta(\hat{\theta}, u_k) \\ &\quad - G'_\theta(\theta_k, u_k))(\theta_{k+1} - \hat{\theta}) - \mathcal{E}_1(\theta_{k+1}, \hat{u}, u_k) + \mathcal{E}_2(\theta_{k+1}, \hat{\theta}, \hat{u}), \end{aligned} \quad (2.23)$$

and takes into account that $\hat{u}$ is a solution to (1.2), then one concludes

$$\begin{aligned} &G_u^{f*}(\theta_{k+1}, u_k)(G(\theta_{k+1}, u_k) - g) + B^{f*}(u_k)(B(u_k) - d) \\ &= G_u^{f*}(\theta_{k+1}, u_k) \left[G'_u(\theta_{k+1}, u_k)(u_k - \hat{u}) + G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta}) \right. \\ &\quad \left. + \mathcal{P}(\theta_k, \theta_{k+1}, \hat{\theta}, u_k, \hat{u}) \right] + B^{f*}(u_k)[B'(u_k)(u_k - \hat{u}) + \mathcal{L}(u_k, \hat{u})] \end{aligned} \quad (2.24)$$

with $\mathcal{P}$ and $\mathcal{L}$ satisfying (2.17). Identities (1.9) and (2.24) imply (2.16), and the proof of lemma 2.2 is complete. Expression (2.16) brings us to the key part of the convergence analysis, which is the estimate of $\|G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta})\|$. This estimate is established in the following lemma.

Lemma 2.3. *Let conditions A, B, (1.11), and (2.1) be met, and $\begin{pmatrix} \theta_k \\ u_k \end{pmatrix}$ and $\begin{pmatrix} \theta_{k+1} \\ u_k \end{pmatrix} \in \mathcal{O}_\eta(\hat{q})$. Then for $\{\theta_k\}$ defined in (1.8), one has*

$$\|G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta})\| \leq \frac{C}{2} \|\theta_k - \hat{\theta}\|^2 + \alpha_k \varepsilon + \frac{C\varepsilon}{2} \sqrt{\frac{\alpha_k}{\lambda}} (\|u_k - \hat{u}\| + \|\theta_k - \hat{\theta}\|). \quad (2.25)$$

Proof. From lemma 2.1, one derives

$$\begin{aligned} G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta}) &= -G'_\theta(\theta_k, u_k)[G'^{*}_\theta(\theta_k, u_k)G'_\theta(\theta_k, u_k) + \alpha_k T^* T]^{-1} G'^{*}_\theta(\theta_k, u_k) \\ &= \{B(\theta_k, \hat{\theta}, u_k, \hat{u}) + \alpha_k v\} - \alpha_k G'_\theta(\theta_k, u_k) [G'^{*}_\theta(\theta_k, u_k) \\ &\quad \times G'_\theta(\theta_k, u_k) + \alpha_k T^* T]^{-1} \{G'_\theta(\hat{\theta}, \hat{u}) - G'_\theta(\theta_k, u_k)\}v, \end{aligned} \quad (2.26)$$

and therefore

$$\begin{aligned} \|G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta})\| &\leq \|G'_\theta(\theta_k, u_k)[G'^{*}_\theta(\theta_k, u_k)G'_\theta(\theta_k, u_k) + \alpha_k T^* T]^{-1} G'^{*}_\theta(\theta_k, u_k)\| \\ &\quad \times \{\|B(\theta_k, \hat{\theta}, u_k, \hat{u})\| + \alpha_k \varepsilon\} + \alpha_k \varepsilon \|G'_\theta(\theta_k, u_k) [G'^{*}_\theta(\theta_k, u_k) \\ &\quad \times G'_\theta(\theta_k, u_k) + \alpha_k T^* T]^{-1} \|G'_\theta(\hat{\theta}, \hat{u}) - G'_\theta(\theta_k, u_k)\|. \end{aligned} \quad (2.27)$$

Using representation (2.11) and the definition of A_k introduced in lemma 2.1, one arrives at the following inequality

$$\begin{aligned} \|G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta})\| &\leq \|(A_k^* A_k)^{1/2} [A_k^* A_k + \alpha_k I]^{-1} (A_k^* A_k)^{1/2}\| \{\|B(\theta_k, \hat{\theta}, u_k, \hat{u})\| \\ &\quad + \alpha_k \varepsilon\} + \alpha_k \varepsilon \|(A_k^* A_k)^{1/2} [A_k^* A_k + \alpha_k I]^{-1} (T^* T)^{-1/2}\|, \\ &\quad \times \|G'_\theta(\hat{\theta}, \hat{u}) - G'_\theta(\theta_k, u_k)\|. \end{aligned} \quad (2.28)$$

Estimate (2.28) together with (1.11), (2.3), and (2.15) yield (2.25), which concludes the proof of lemma 2.3.

Summarizing the above results, we prove the last lemma of this section, which provides the upper bound for $\|u_k - \hat{u}\|$.

Lemma 2.4. *Let assumptions of lemmas 2.1–2.3 be fulfilled. Then*

$$\begin{aligned} \|u_{k+1} - \hat{u}\| &\leq \frac{C\sqrt{s}}{2} \|\theta_k - \hat{\theta}\|^2 + \frac{3C\sqrt{s}}{4} \|u_k - \hat{u}\|^2 + \frac{C\sqrt{s}}{2} \|\theta_{k+1} - \hat{\theta}\|^2 \\ &\quad + C\sqrt{s}(\|\theta_k - \hat{\theta}\| + \|u_k - \hat{u}\|) \|\theta_{k+1} - \hat{\theta}\| \\ &\quad + \frac{C\varepsilon}{2} \sqrt{\frac{s\alpha_k}{\lambda}} (\|u_k - \hat{u}\| + \|\theta_k - \hat{\theta}\|) + \sqrt{s}\alpha_k \varepsilon. \end{aligned} \quad (2.29)$$

Proof. In light of lemma 2.2, one has

$$\begin{aligned} \|u_{k+1} - \hat{u}\| &= \|[G'^{*}_u(\theta_{k+1}, u_k)G'_u(\theta_{k+1}, u_k) + B'^{*}(u_k)B'(u_k)]^{-1} G'^{*}_u(\theta_{k+1}, u_k)\| \\ &\quad \times \left\{ \|G'_\theta(\theta_k, u_k)(\theta_{k+1} - \hat{\theta})\| + \|\mathcal{P}(\theta_k, \theta_{k+1}, \hat{\theta}, u_k, \hat{u})\| \right\} \\ &\quad + \|[G'^{*}_u(\theta_{k+1}, u_k)G'_u(\theta_{k+1}, u_k) \\ &\quad + B'^{*}(u_k)B'(u_k)]^{-1} B'^{*}(u_k)\| \|\mathcal{L}(u_k, \hat{u})\|, \end{aligned} \quad (2.30)$$

Conditions (2.11) and (1.13) imply

$$\begin{aligned} &\|[G'^{*}_u(\theta_{k+1}, u_k)G'_u(\theta_{k+1}, u_k) + B'^{*}(u_k)B'(u_k)]^{-1} G'^{*}_u(\theta_{k+1}, u_k)\| \\ &= \|[G'^{*}_u(\theta_{k+1}, u_k)G'_u(\theta_{k+1}, u_k)]^{-1/2} [C_k^* C_k + I]^{-1} \\ &\quad \times (G'^{*}_u(\theta_{k+1}, u_k)G'_u(\theta_{k+1}, u_k))^{-1/2} G'^{*}_u(\theta_{k+1}, u_k)\|, \end{aligned}$$

where $C_k := B'(u_k)(G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k))^{-1/2}$. Thus applying polar decomposition to the linear operator $G_u'(\theta_{k+1}, u_k)$, one obtains

$$\begin{aligned} & \| [G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k) + B'^*(u_k)B'(u_k)]^{-1}G_u'^*(\theta_{k+1}, u_k) \| \\ & \leq \| (G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k))^{-1/2}[C_k^*C_k + I]^{-1} \| \leq \sqrt{s} \sup_{t \in \sigma(C_k^*C_k)} \frac{1}{t+1} = \sqrt{s}, \end{aligned} \quad (2.31)$$

$\sigma(C_k^*C_k) \subseteq [0, \|C_k\|^2]$. Furthermore,

$$\begin{aligned} & \| [G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k) + B'^*(u_k)B'(u_k)]^{-1}B'^*(u_k) \| \\ & = \| (G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k))^{-1/2}[C_k^*C_k + I]^{-1} \\ & \quad (G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k))^{-1/2}B'^*(u_k) \| \\ & = \| (G_u'^*(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k))^{-1/2}[C_k^*C_k + I]^{-1}C_k^* \| \\ & \leq \sqrt{s} \sup_{t \in \sigma(C_k^*C_k)} \frac{\sqrt{t}}{t+1} = \frac{\sqrt{s}}{2}. \end{aligned} \quad (2.32)$$

Combining (2.25), (2.30)–(2.32), one arrives at (2.29). Thus estimate (2.29) is true as claimed.

2.2. Convergence analysis in the noise-free case

In the next two subsections, we use the above four lemmas to establish our main convergence results, theorems 2.5 and 2.7. To that end, we prove by induction that for θ_0 and u_0 sufficiently close to $\hat{\theta}$ and $\hat{u}$, respectively, there are positive constants, l_1 and l_2 , such that

$$\|\theta_k - \hat{\theta}\| \leq l_1\sqrt{\alpha_k} \quad \text{and} \quad \|u_k - \hat{u}\| \leq l_2\alpha_k, \quad k = 1, 2, \dots \quad (2.33)$$

Indeed, assume that θ_0 and u_0 are chosen to satisfy (2.33) when $k = 0$, and let (2.33) be fulfilled for any $j = 1, 2, \dots, k$. Then by (1.18), (2.4) and by induction assumption, one concludes

$$\begin{aligned} \|\theta_{k+1} - \hat{\theta}\| & \leq \frac{1}{2\sqrt{\lambda\alpha_k}} \left[C\|u_k - \hat{u}\| + \frac{C}{2}\|\theta_k - \hat{\theta}\|^2 + \alpha_k\varepsilon \right] + \frac{C\varepsilon}{\lambda}(\|u_k - \hat{u}\| + \|\theta_k - \hat{\theta}\|) \\ & \leq \left\{ \frac{a}{2\sqrt{\lambda}} \left[Cl_2 + \frac{C}{2}l_1^2 + \alpha_k\varepsilon \right] + \frac{C\varepsilon a}{\lambda}(l_2\sqrt{\alpha_0} + l_1) \right\} \sqrt{\alpha_{k+1}}. \end{aligned} \quad (2.34)$$

To ensure that the right-hand side of (2.34) does not exceed $l_1\sqrt{\alpha_{k+1}}$, one has to select l_1 in such a way that

$$\frac{Ca}{4\sqrt{\lambda}}l_1^2 - \left[1 - \frac{C\varepsilon a}{\lambda} \right] l_1 + a \left[Cl_2 \left(\frac{1}{2\sqrt{\lambda}} + \frac{\varepsilon\sqrt{\alpha_0}}{\lambda} \right) + \frac{\varepsilon}{2\sqrt{\lambda}} \right] \leq 0.$$

For example, one can take

$$l_1 := \frac{2(\lambda - C\varepsilon a)}{\sqrt{\lambda}Ca} \quad (2.35)$$

as long as

$$\frac{\lambda}{Ca} \geq \varepsilon + \sqrt{\lambda \left[\left(\frac{1}{2} + \varepsilon\sqrt{\frac{\alpha_0}{\lambda}} \right) l_2 + \frac{\varepsilon}{2C} \right]}. \quad (2.36)$$

Likewise, from (2.29), it follows that for $\|u_{k+1} - \hat{u}\|$ to be less than $l_2\alpha_{k+1}$, the constant l_2 must satisfy the inequality

$$\begin{aligned} & \frac{3C\sqrt{s}\alpha_0a^2}{4}l_2^2 - \left[1 - C\sqrt{s\alpha_0}a\left(l_1 + \frac{\varepsilon a}{2\sqrt{\lambda}}\right)\right]l_2 + \frac{C\sqrt{s}}{2}(a+1)^2l_1^2 \\ & + \frac{C\sqrt{s}a^2\varepsilon}{2\sqrt{\lambda}}l_1 + \sqrt{s}\varepsilon a^2 \leq 0. \end{aligned} \quad (2.37)$$

Substituting (2.35) into (2.37), one obtains

$$\begin{aligned} & \frac{3C\sqrt{s}\alpha_0a^2}{4}l_2^2 - \left[1 - \sqrt{\frac{\alpha_0s}{\lambda}}\left(2(\lambda - C\varepsilon a) + \frac{\varepsilon a^2C}{2}\right)\right]l_2 \\ & + \frac{2\sqrt{s}(a+1)^2(\lambda - C\varepsilon a)^2}{\lambda Ca^2} + \frac{\sqrt{s}a\varepsilon(\lambda - C\varepsilon a)}{\lambda} + \sqrt{s}\varepsilon a^2 \leq 0. \end{aligned} \quad (2.38)$$

Condition (2.38) is fulfilled when

$$l_2 := \frac{\frac{2}{\sqrt{s}} - \sqrt{\frac{\alpha_0}{\lambda}}(4(\lambda - C\varepsilon a) + \varepsilon a^2C)}{3C\alpha_0a^2}, \quad (2.39)$$

and the following inequality holds

$$\sqrt{\frac{\lambda}{s\alpha_0}} \geq 2(\lambda - C\varepsilon a) + \frac{\varepsilon a^2C}{2} + \sqrt{3\{2(a+1)^2(\lambda - C\varepsilon a)^2 + Ca^3\varepsilon(\lambda - C\varepsilon a) + Ca^4\lambda\varepsilon\}}. \quad (2.40)$$

Finally, plugging in the value of l_2 into estimate (2.36), one derives

$$\frac{\sqrt{\lambda}}{C} \geq \frac{\varepsilon a}{\sqrt{\lambda}} + \sqrt{\frac{(1 + 2\varepsilon\sqrt{\frac{\alpha_0}{\lambda}})\left\{\frac{2}{\sqrt{s}} - \sqrt{\frac{\alpha_0}{\lambda}}(4(\lambda - C\varepsilon a) + \varepsilon a^2C)\right\} + 3\varepsilon\alpha_0a^2}{6C\alpha_0}}. \quad (2.41)$$

Summarizing the above, we arrive at the following theorem.

Theorem 2.5 (Noise-free case). *Let conditions of lemmas 2.1–2.4 be satisfied with*

$$\eta := \sqrt{l_1^2\alpha_0 + l_2^2\alpha_0^2} \quad (2.42)$$

and with $l_1 > 0$ and $l_2 > 0$ defined in (2.35) and (2.39), respectively. Suppose that regularization sequence, $\{\alpha_k\}$, is chosen according to (1.8) and

$$\|\theta_0 - \hat{\theta}\| \leq l_1\sqrt{\alpha_0} \quad \text{and} \quad \|u_0 - \hat{u}\| \leq l_2\alpha_0. \quad (2.43)$$

Assume that λ in (1.1), ε in (2.1), and α_0 and a in (1.8) are such that inequalities (2.40) and (2.41) hold. Then in the noise-free case, for $\{\theta_k\}$, generated by (1.8), and for $\{u_k\}$, generated by (1.9), estimates (2.33) are fulfilled.

Remark 2.6. Note that inequality (2.40) is satisfied when $\lambda - C\varepsilon a$ and ε are small, and inequality (2.41) will always be true for sufficiently large α_0 .

2.3. Convergence rates in case of noise-contaminated data

Now consider the case of noisy data. That is, suppose that in (1.2), $\|d - d_\delta\| \leq \delta$ with $\delta > 0$. Then (2.16) takes the form

$$\begin{aligned} u_{k+1} - \hat{u} = & -[G_u^{t*}(\theta_{k+1}, u_k)G_u'(\theta_{k+1}, u_k) + B^{t*}(u_k)B'(u_k)]^{-1} \\ & \times \left\{ G_u^{t*}(\theta_{k+1}, u_k)[G_\theta'(\theta_k, u_k)(\theta_{k+1} - \hat{\theta}) + \mathcal{P}(\theta_k, \theta_{k+1}, \hat{\theta}, u_k, \hat{u})] \right. \\ & \left. + B^{t*}(u_k)(\mathcal{L}(u_k, \hat{u}) + d - d_\delta) \right\}. \end{aligned} \quad (2.44)$$

Let δ be chosen in such a way that

$$\delta \leq \nu\alpha_0, \quad \text{for some } \nu > 0, \quad (2.45)$$

and let iterations (1.8) and (1.9) be terminated when $k = \mathcal{K}(\delta)$, where

$$\mathcal{K}(\delta) = \max\{k = 0, 1, 2, \dots : \delta \leq \nu\alpha_{k-1}\}. \quad (2.46)$$

Then for any $k \leq \mathcal{K}(\delta)$, estimate (2.29) becomes

$$\begin{aligned} \|u_{k+1} - \hat{u}\| \leq & \frac{C\sqrt{s}}{2}\|\theta_k - \hat{\theta}\|^2 + \frac{3C\sqrt{s}}{4}\|u_k - \hat{u}\|^2 + \frac{C\sqrt{s}}{2}\|\theta_{k+1} - \hat{\theta}\|^2 \\ & + C\sqrt{s}(\|\theta_k - \hat{\theta}\| + \|u_k - \hat{u}\|)\|\theta_{k+1} - \hat{\theta}\| \\ & + \frac{C\varepsilon}{2}\sqrt{\frac{s\alpha_k}{\lambda}}(\|u_k - \hat{u}\| + \|\theta_k - \hat{\theta}\|) + \sqrt{s}\alpha_k\left(\varepsilon + \frac{\nu}{2}\right) \end{aligned} \quad (2.47)$$

and in place of (2.38) one gets

$$\begin{aligned} & \frac{3C\sqrt{s}\alpha_0a^2}{4}l_2^2 - \left[1 - \sqrt{\frac{\alpha_0s}{\lambda}}\left(2(\lambda - C\varepsilon a) + \frac{\varepsilon a^2C}{2}\right)\right]l_2 \\ & + \frac{2\sqrt{s}(a+1)^2(\lambda - C\varepsilon a)^2}{\lambda Ca^2} + \frac{\sqrt{s}a\varepsilon(\lambda - C\varepsilon a)}{\lambda} + \sqrt{s}\left(\varepsilon + \frac{\nu}{2}\right)a^2 \leq 0. \end{aligned} \quad (2.48)$$

As the result, we state the following theorem.

Theorem 2.7 (The case of noise in the data). Suppose all conditions of theorem 2.5 are met with (2.40) replaced with

$$\begin{aligned} \sqrt{\frac{\lambda}{s\alpha_0}} \geq & 2(\lambda - C\varepsilon a) + \frac{\varepsilon a^2C}{2} \\ & + \sqrt{3\left\{2(a+1)^2(\lambda - C\varepsilon a)^2 + Ca^3\varepsilon(\lambda - C\varepsilon a) + Ca^4\lambda\left(\varepsilon + \frac{\nu}{2}\right)\right\}}, \end{aligned}$$

and suppose that in (1.2), $\|d - d_\delta\| \leq \delta$ with $\delta > 0$. Then if assumptions (2.45) and (2.46) are fulfilled, one has

$$\|\theta_k - \hat{\theta}\| \leq l_1\sqrt{\alpha_k} \quad \text{and} \quad \|u_k - \hat{u}\| \leq l_2\alpha_k, \quad k = 1, 2, \dots, \mathcal{K}(\delta), \quad (2.49)$$

the number $\mathcal{K}(\delta)$ is admissible, i.e., $\lim_{\delta \rightarrow 0} \mathcal{K}(\delta) = \infty$, and the following convergence rates hold

$$\|\theta_{\mathcal{K}(\delta)} - \hat{\theta}\| = O(\sqrt{\delta}) \quad \text{and} \quad \|u_{\mathcal{K}(\delta)} - \hat{u}\| = O(\delta). \quad (2.50)$$

3. Numerical study of inverse problems in epidemiology

3.1. Parameter estimation and forecasting from modified Richards equation

In this section we illustrate advantages and limitations of our proposed method (1.8) and (1.9) using parameter estimation inverse problems in epidemiology. All experiments are carried out with real data on COVID-19 pandemic provided by Dr. Gerardo Chowell from Georgia State University School of Public Health. In our first study, we employ the modified Richards (MR) epidemic model, based on four parameters, p , K , a , and r [30, 31]:

$$\frac{dC}{dt} = rC^p \left[1 - \left(\frac{C}{K} \right)^a \right]. \quad (3.1)$$

These four parameters, characterizing the spread of the virus, are reconstructed numerically by the new PCA. Once recovered, they are used to quantify the natural capacity of the disease and to forecast future incidence cases.

Over the past few months, several compartmental models have been used to study the spread of COVID-19 virus and to examine the impact of control and prevention [2, 12, 28]. These models are based on differential equations that track progression of individuals between different epidemiological states or risk levels. While compartmental models remain the kernel of rich epidemic theory [2, 12], a critical underlying assumption is that random mixing governs the interaction between individuals in population. Consequently, these models exhibit exponential-growth during the early epidemic phase in the absence of significant depletion of susceptible individuals and intervention measures [7]. However, recent findings revealed a diversity of early epidemic growth profiles across infectious diseases, highlighting the presence of sub-exponential and higher than exponential growth, which contrast with traditional assumption of exponential epidemic spread [7, 9]. In case of COVID-19, substantial heterogeneity in susceptibility and infectivity of the host population can significantly distort the contact network structure [7, 9]. Thus, epidemic models that capture various growth rates during the first few generations of the outbreak are needed to better characterize the COVID-19 transmission dynamics. For this reason, in order to assess an early stage of COVID-19, we use phenomenological model (3.1). By its very design, this modified logistic equation, originating from the field of ecology, is able to capture a diversity of early growth profiles for a variety of epidemics, leading to more accurate parameter estimation and more reliable forecasting trajectories [9, 30, 31].

In (3.1), $C(t)$ and $\frac{dC}{dt}$ are cumulative number of cases and the number of new incidence cases at time t , respectively. The deceleration of growth parameter, p , modulates different growth patterns in case incidence [9], r is the intrinsic growth rate, a measures the extent of deviation from the S-shaped dynamics of the classical logistic growth model [30], and K represents the epidemic final size, defined as the total number of infections throughout the outbreak. When $p = 1$, (3.1) is known as the Richards model [31] and its analytical solution for the cumulative number of cases is given by

$$C(t) = \frac{K}{[1 + a e^{-ar(t-\tau)}]^{1/a}}, \quad (3.2)$$

with τ being the inflection point of C . If $p \neq 1$, the solution to (3.1) is an infinite series [30], which cannot be written in the closed form. If one differentiates both sides of (3.1) with respect to t and equates them to zero, then one obtains the cumulative number of cases at the disease

inflection point, τ :

$$C(\tau) = K \left(\frac{p}{a+p} \right)^{\frac{1}{a}}. \quad (3.3)$$

Substituting (3.3) into the right-hand side of (3.1), one derives

$$C'(\tau) = \frac{raK^p}{p} \left(\frac{p}{a+p} \right)^{1+\frac{p}{a}},$$

the maximum incidence in the modified Richards model. Estimation of K , $C(\tau)$, and $C'(\tau)$ for an emerging disease provides crucial information needed for adequate intervention and control measures to be put in place. Assume that a wave of an outbreak originates on day (week) t_1 and ends on day (week) $\mathcal{B}$ (which is unknown at the early stage of disease transmission). Let the incidence data, d_δ , be reported at $t = t_1 < t_2 < \dots < t_n$ with t_n being much smaller than $\mathcal{B}$. In that case, given limited data, d_δ , at the start of a new wave one has to solve the following ODE-constrained minimization problem

$$\min_C \left\| \frac{dC}{dt} - d_\delta \right\|^2 = \min_C \sum_{j=1}^n \left(\frac{dC}{dt}(t_j) - d_\delta(t_j) \right)^2, \quad (3.4)$$

$$\text{subject to } \frac{dC}{dt} = rC^p \left[1 - \left(\frac{C}{K} \right)^a \right], \quad C(t_1) = C^{(1)}. \quad (3.5)$$

It has been observed in [25], that estimation of the unknown parameters, p , r , a , and K , from problem (3.4) and (3.5) in a ‘traditional way’ given early incidence data, d_δ , that is, minimizing $\left\| \frac{dC}{dt} - d_\delta \right\|^2$ with respect to p , r , a , and K while solving equation (3.5) numerically at every step of the iterative process, is extremely unstable in K , leading to very inaccurate predictions of disease capacity. Therefore, in this subsection, we use our newly proposed PCA (1.8) and (1.9) in order to approximate a solution to (3.4) and (3.5) in a stable manner for $t \in [t_1, t_n]$ with $t_n < \mathcal{B}$.

Suppose that $t = \mathcal{A}$ is one day (week) before the first case is reported, i.e., $C(\mathcal{A}) = C'(\mathcal{A}) = 0$. Denote $E(t) := \frac{dC}{dt}$ and discretize $E(t)$ using Fourier approximation:

$$E(t) = A_0 + \sum_{j=1}^N \left\{ A_j \cos \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) + B_j \sin \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) \right\}. \quad (3.6)$$

To ensure that $E(\mathcal{A}) = E(\mathcal{B}) = 0$, where $\mathcal{B}$ is an estimated last day (week) of the current epidemic wave, we set $A_0 = -\sum_{j=1}^N A_j$, and conclude

$$E(t) = \sum_{j=1}^N \left\{ A_j \left[\cos \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right] + B_j \sin \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) \right\}. \quad (3.7)$$

Then one obtains

$$\begin{aligned} C(t) = \int_{\mathcal{A}}^t E(s) ds &= \sum_{j=1}^N \left\{ A_j \left[\frac{\mathcal{B} - \mathcal{A}}{2\pi j} \sin \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - (t - \mathcal{A}) \right] \right. \\ &\quad \left. - B_j \frac{\mathcal{B} - \mathcal{A}}{2\pi j} \left[\cos \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right] \right\}. \end{aligned} \quad (3.8)$$

Introduce the notations:

$$u := [A_1, \dots, A_N, B_1, \dots, B_N]^T, \quad \text{and} \quad \theta_1 := p, \quad \theta_2 := K, \quad \theta_3 := a, \quad \text{and} \quad \theta_4 := r. \quad (3.9)$$

The discrete approximation of $E(t)$ allows us to define $B(u)$ in (1.3) as a linear operator from $\mathbb{R}^{2N}$ to $\mathbb{R}^n$ (with n being the number of data points) such that

$$B_i(u) := \sum_{j=1}^N \left\{ u_j \left[\cos \left(2\pi j \frac{t_i - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right] + u_{N+j} \sin \left(2\pi j \frac{t_i - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) \right\}, \quad i = 1, 2, \dots, n. \quad (3.10)$$

Furthermore, based on (3.5), in minimization problem (1.3) we define

$$G_i(\theta, u) := B_i(u) - \theta_4 C_i^{\theta_1}(u) \left[1 - \left(\frac{C_i(u)}{\theta_2} \right)^{\theta_3} \right], \quad G: \mathbb{R}^4 \times \mathbb{R}^{2N} \rightarrow \mathbb{R}^n, \quad \text{and} \quad g := 0, \quad (3.11)$$

where

$$C_i(u) = \sum_{j=1}^N \left\{ u_j \left[\frac{\mathcal{B} - \mathcal{A}}{2\pi j} \sin \left(2\pi j \frac{t_i - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - (t_i - \mathcal{A}) \right] + u_{N+j} \frac{\mathcal{B} - \mathcal{A}}{2\pi j} \left[\cos \left(2\pi j \frac{t_i - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right] \right\}.$$

Expressions (3.10) and (3.11) enable us to easily calculate Fréchet derivatives of $B(u)$ and $G(\theta, u)$ with respect to u and the derivative of $G(\theta, u)$ with respect to θ , and to implement algorithm (1.8) and (1.9) for estimating $\theta = [p, K, a, r]^T$ from (3.4) and (3.5). Given a considerable gap in the levels of magnitude between K and the rest of the parameters, in all our simulations T^*T in (1.8) is a diagonal matrix with entries selected to scale initial values of the unknowns [27]:

$$T^*T = \begin{bmatrix} w^2 & 0 & 0 & 0 \\ 0 & w_K^2 & 0 & 0 \\ 0 & 0 & w^2 & 0 \\ 0 & 0 & 0 & w^2 \end{bmatrix}. \quad (3.12)$$

Without scaling (that is, with $T = I$) the process turns out to be divergent for all initial approximations tested. To study the efficiency of predictor–corrector scheme (1.8) and (1.9), we take COVID-19 incidence case data for Netherlands and Spain, two countries with different size of the population (17 134 872 and 46 754 778 people in 2020, respectively) and different scope of the pandemic (47 335 and 283 941 total cases as of June 7, 2020, respectively).

As our first experiment, we reconstruct u and θ from full incidence data set for Netherlands. As full data set, we consider COVID-19 incidence data from February 28, 2020, when the first case was reported in Netherlands, till May 24, 2020, the time this experiment was conducted (87 data points). To initiate the algorithm, we assume $p_0 = a_0 = 1$, $r_0 = 0.1$, and $K_0 = 60\,000$. To get the initial guess for u , we solve (3.1) analytically on the interval $[1, 87]$ with $p = a = 1$, $r = 0.5$, $K = 60\,000$, and $C(1) = 1$ (the number of cases reported on February 28, 2020), and then (numerically) calculate the Fourier series expansion coefficients for $rC^p \left[1 - \left(\frac{C}{K} \right)^a \right]$.

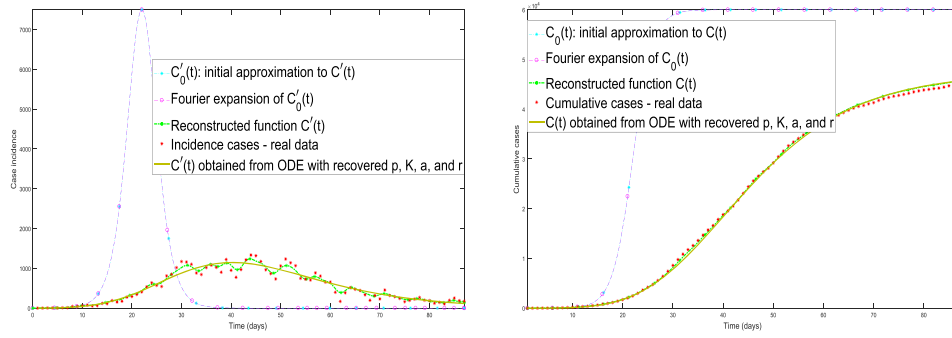


Figure 1. Netherlands: reconstructed incidence and cumulative cases from a full data set.

The initial values of $C(t)$ and $C'(t)$ together with their Fourier approximations are illustrated in figure 1. Even though $C_0(t)$ and $C'_0(t)$ are far from accurate, algorithm (1.8) and (1.9) is still convergent for a suitable choice of control elements. As $\bar{\theta}$, we take $[0.550\ 0000\ 50.5]^T$. In (3.7), for the full data set N is equal to 30 (i.e., the number of base functions is 60), though a very similar result can be obtained with a much smaller N . The value of $\mathcal{B}$ is estimated as 100.

Since we expect K and the rest of the parameters to be about 5 or 6 orders of magnitude apart, in (3.12) we take $w_K = 0.32$ and $w = 10^4$. The regularization sequence is chosen to be $\alpha_k = \alpha_0/k^4$ with $\alpha_0 = 1$ to ensure the most aggressive convergence rate that can be achieved while still keeping the process stable. The iterations are stopped when $k = 15$ and $\alpha = 1.98 \times 10^{-5}$. This stopping time is determined by the goodness of fit (with no overfitting) to the data for both, the reconstructed $C(t)$ and $C'(t)$ and for $C(t)$ and $C'(t)$ obtained from (3.1) with recovered parameters p , K , a , and r . For Netherlands full data set, we obtain

$$p = 0.830\ 30, \quad K = 47\ 351, \quad a = 0.526\ 96, \quad \text{and} \quad r = 0.841\ 20.$$

The next phase of our experiment is to forecast future incidence cases given 20, 30, and 50 data points for Netherlands. After p , K , a , and r have been recovered from early epidemic data, 100 additional bootstrap curves are generated by adding Poisson error structure to the daily series of reported cases [8, 10, 11] in order to quantify uncertainty in the recovered parameters. The histograms in the upper row (figures 2–4) show the approximate values for the components of θ along with the 95% confidence intervals, while the collection of curves at the bottom of the figure demonstrates the accuracy of forecasting. The forecasting curves are computed by solving ODE (3.1) on the entire interval $[\mathcal{A}, \mathcal{B}]$ with parameters p , K , a , and r recovered from early available data using algorithm (1.8) and (1.9). The brown solid vertical line in figures 2–4 separates the calibration and forecasting periods.

In figure 2, one can see forecasting results for Netherlands from the first 20 incidence data points. While the forecasting curves show a very accurate turning time, the number of cases is clearly underestimated. The uncertainty of future projections is rather high as expected for such an early prognosis. To recover θ from limited data, we use $p_0 = a_0 = r_0 = 1$, and $K_0 = 60\ 000$. To get the initial guess for u , we solve (3.1) analytically with $p = a = 1$, $r = 0.5$, $K = 60\ 000$, and $C(1) = 1$ and then calculate the Fourier series expansion coefficients for $rC^p [1 - (\frac{C}{K})^a]$. We take $\bar{\theta} = [0.5\ 20\ 000\ 0.5\ 1.5]^T$. As extra regularization tool, we reduce the size of the solution space for u from 60 ($N = 30$ in (3.7)) to 10 ($N = 5$), since the process becomes increasingly more unstable with less data. The regularization sequence is still $\alpha_k = \alpha_0/k^4$, and α_0 is chosen

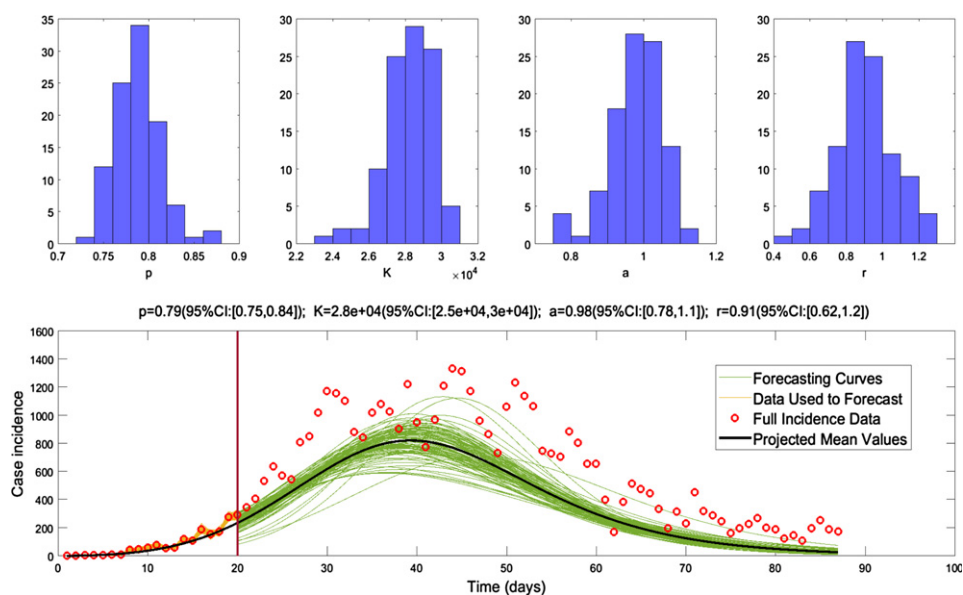


Figure 2. Netherlands: forecasting from 20 days of incidence data.

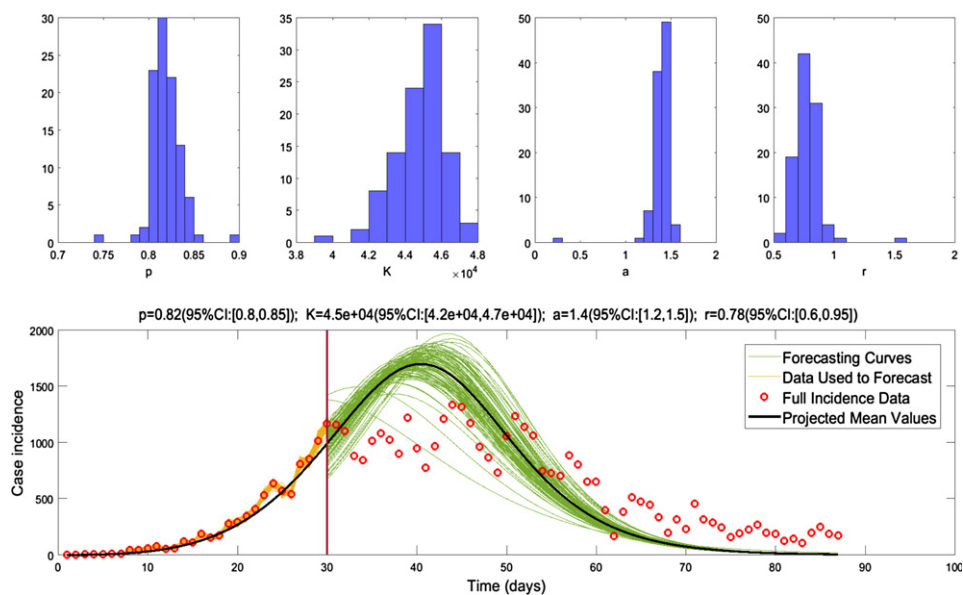


Figure 3. Netherlands: forecasting from 30 days of incidence data.

to be 2×10^{-2} . The iterations are stopped when $k = 9$ and $\alpha_k = 8.33 \times 10^{-6}$. This stopping time is determined by the goodness of fit to the first 20 data points.

Forecasting curves for 30 data points (figure 3) show that 10 extra days can make a big difference and it is worth recomputing future projections every day (week) if forecasting is

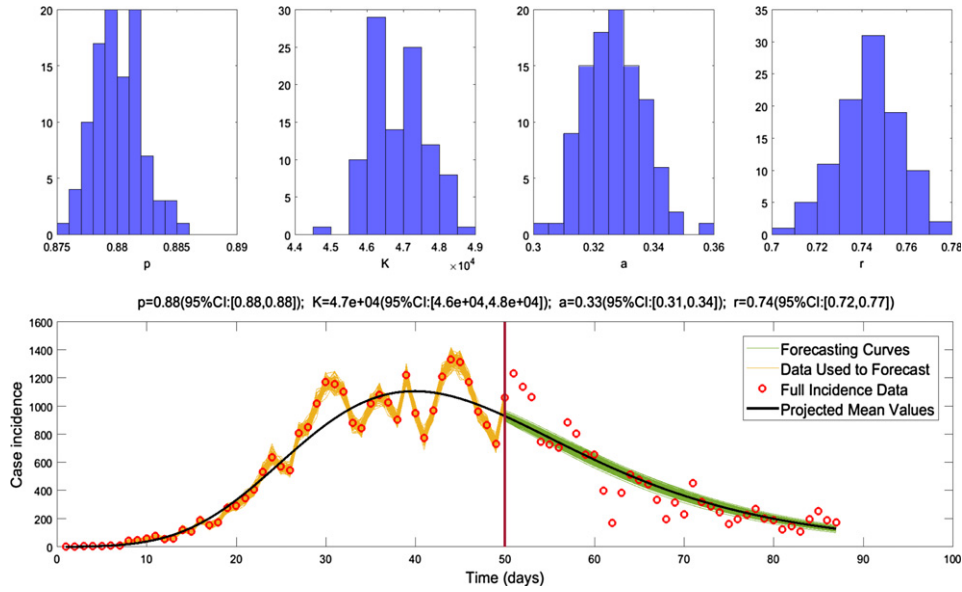


Figure 4. Netherlands: forecasting from 50 days of incidence data.

done in real time. For the first 20 days of the forecasting period, the daily number of new cases is overestimated, which is not a surprise given the sharp upward trend in the available data at the end of the calibration window. The next two weeks of projected data (from day 50 to 62) are adequately covered by the bundle of forecasting curves. For this experiment, we take the same initial values for p , K , a , and r as before and compute u_0 in the same manner. The number of base functions used to discretize $C'(t)$ is also the same, 10. The regularization sequence, $\alpha_k = \alpha_0/k^4$, while $\alpha_0 = 1$ and iterations are stopped when $k = 14$ ($\alpha_k = 2.60 \times 10^{-5}$).

Forecasting after the turning time (figure 4) is less useful from practical standpoint, but it is still important as we examine the efficiency of the algorithm. With 50 data points, we set $p_0 = a_0 = 1$, $r_0 = 0.1$, $K_0 = 60\,000$, and $\bar{\theta} = [0.5 \ 50\,000 \ 0.5 \ 0.5]^T$. The regularization sequence $\alpha_k = \alpha_0/k^4$ with $\alpha_0 = 50$. The iterations are stopped when $k = 17$ and $\alpha_k = 5.99 \times 10^{-4}$. This stopping time is determined by the goodness of fit to the data used for training (the first 50 data points). As expected, the forecasting results in this case are accurate with low uncertainty.

In figure 5, one can see the results of numerical simulations with full data set for Spain. This data set covers the period from February 20, 2020, the day when the first case was reported in Spain, till June 7, 2020, the time this experiment was carried out (109 data points). We assume that $\mathcal{B}$, the length of the epidemic wave, is equal to 120. To initiate the algorithm, we take $p_0 = a_0 = 1$, $r_0 = 0.1$, and $K_0 = 200\,000$. In order to obtain the initial guess for u , we solve (3.1) analytically on $[1, 109]$ with $p = a = 1$, $r = 0.5$, $K = 200\,000$, and $C(1) = 2$ (the number of cases reported on February 20, 2020), and then calculate the Fourier series expansion coefficients for $rC^p [1 - (\frac{C}{K})^a]$. The initial values for $C(t)$ and $C'(t)$ together with their Fourier approximations are illustrated in figure 6. Clearly, these initial values are not close to the solution, and nevertheless algorithm (1.8) and (1.9) is convergent for a broad range of control elements. The reference vector, $\bar{\theta}$, is chosen to be $[0.5300\,0000.50.5]^T$, while the number of base functions for the full data set is 60 (i.e., $N = 30$ in (3.7)), the same as in the case of Netherlands. With fewer base functions, the results of the experiment are very similar.

For Spain, we expect K and the rest of the parameters to be about 6 or 7 orders of magnitude apart. Thus in (3.12), we take $w_K = 0.1$ and $w = 31\,623$. The regularization sequence is chosen

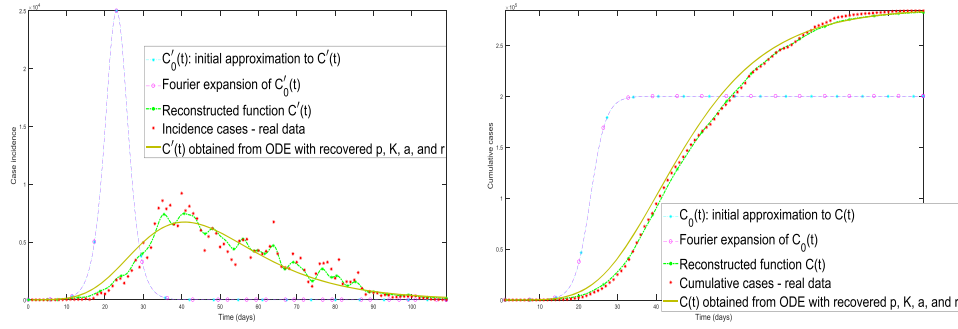


Figure 5. Spain: reconstructed incidence and cumulative cases from a full data set.

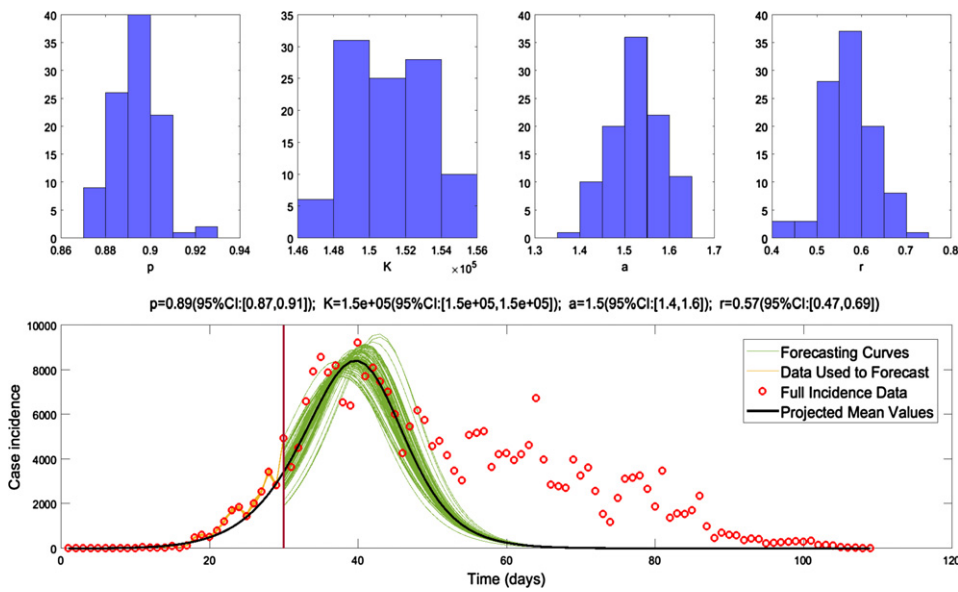


Figure 6. Spain: forecasting from 30 days of incidence data.

to be $\alpha_k = \alpha_0/k^3$ with $\alpha_0 = 30$ to accelerate convergence without jeopardizing stability. The iterations are stopped when $k = 15$ and $\alpha_k = 0.0089$. This stopping time is determined by the goodness of fit to the data for both, the reconstructed $C(t)$ and $C'(t)$ and for $C(t)$ and $C'(t)$ obtained from (3.1) with recovered parameters p , K , a , and r . For the full set of COVID-19 data in Spain, we get

$$p = 0.91378, \quad K = 286559, \quad a = 0.21366, \quad \text{and} \quad r = 0.89943.$$

Figures 6 and 7 illustrate histograms with parameter values and the bundles of curves showing projected case incidence given 30 and 45 days of COVID-19 data in Spain. For every partial data set, 100 additional bootstrap curves are generated by adding Poisson error structure to the daily series of reported cases [10] in order to quantify uncertainty in the approximate values of the parameters. The forecasting curves are computed by solving ODE (3.1) on the entire

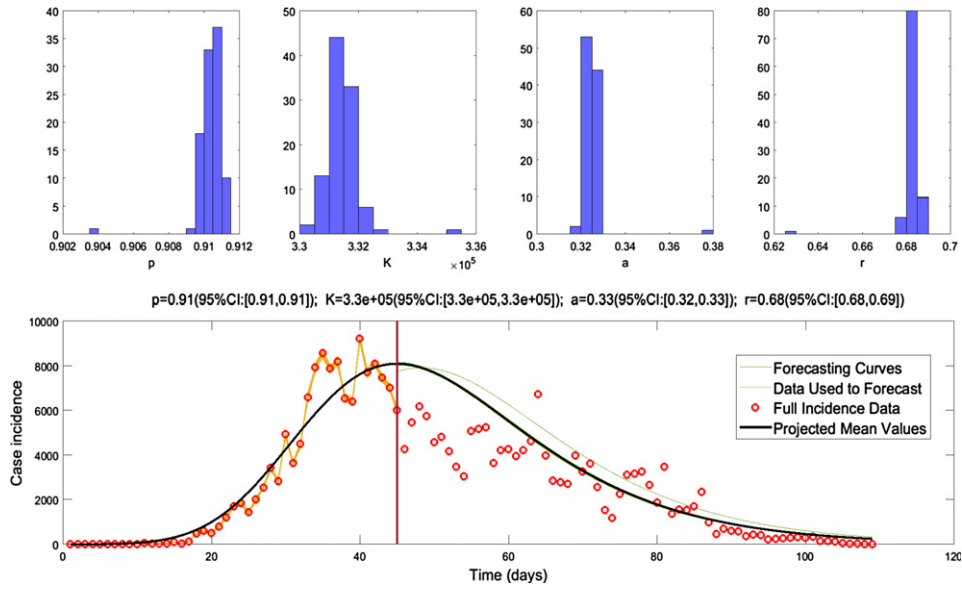


Figure 7. Spain: forecasting from 45 days of incidence data.

interval $[A, B]$ with parameters p , K , a , and r recovered from each bootstrap curve by algorithm (1.8) and (1.9).

In figure 6, forecasting results for Spain with 30 data points are given. The projected case incidence is remarkably accurate for the first 20 days of forecasting interval. The forecasting curve bundle does not, however, cover the rest of the data showing a clear need to recompute the anticipated values as more information becomes available. The uncertainty of future projections is rather high, but reasonable, considering a relatively short calibration period. For parameter estimation, we use $p_0 = a_0 = 1$, $r_0 = 0.1$, and $K_0 = 100\,000$. To obtain u_0 , we solve (3.1) analytically with $p = a = 1$, $r = 0.5$, $K = 100\,000$, and $C(1) = 2$ and then calculate the Fourier series expansion coefficients for $rC^p \left[1 - \left(\frac{C}{K}\right)^a\right]$. As we move from full to partial data, the number of base functions for $C(t)$ and $C'(t)$ is reduced from 60 ($N = 30$ in (3.7)) to 10 ($N = 5$) (similar to how it was done for Netherlands), since the process becomes increasingly more unstable with less data. The regularization sequence $\alpha_k = 1/k^4$, and the reference vector $\bar{\theta} = [0.5 \ 50\,000 \ 0.5 \ 0.5]^T$. The iterations are stopped when $k = 15$ and $\alpha_k = 1.98 \times 10^{-5}$. This stopping time is determined by the goodness of fit to the first 30 data points.

In figure 7, one can see parameter values and future projections of COVID-19 incidence cases in Spain, estimated from 45 data points. The figure illustrates forecasting results at the very pick of the incidence curve. The forecasting bundles do not account for a sharp unexpected drop in incidence cases following the end of the calibration period. But in a long run, the prognosis is accurate with virtually no uncertainty. With 45 data points, we set $p_0 = a_0 = 1$, $r_0 = 2$, $K_0 = 500\,000$, and $\bar{\theta} = [0.5 \ 350\,000 \ 0.5 \ 0.5]^T$. The Fourier coefficients comprising u_0 get calculated from the analytic solution to (3.1) with $p = a = r = 1$, and $K = 500\,000$. The regularization sequence $\alpha_k = \alpha_0/k^3$ with $\alpha_0 = 10$. The iterations are stopped when $k = 10$ and $\alpha_k = 0.01$. This stopping time is determined by the goodness of fit to the first 45 data points used for training.

3.2. Application of SIRD model

Apart from forecasting curves with quantified uncertainty, the reproductive capacity of an outbreak and the underlying transmission rate are important tools, which allow to assess the efficiency of public health policies. Whereas other system parameters, i.e. incubation and recovery rates, are less dependent on intervention measures, the effective reproduction number and the transmission rate of a disease are directly influenced by control and prevention. Having the tools needed to recover these two parameters (call them $\mathcal{R}(t)$ and $\beta(t)$, respectively) allows for the real-time analysis of the effectiveness of intervention measures, for the ability to determine the most powerful response and, finally, for the conceivably more accurate projections of future incidence cases.

In this subsection we present our numerical results on stable estimation of $\mathcal{R}(t)$ and $\beta(t)$ from COVID-19 incidence data using a version of SIRD model introduced in [2]:

$$\frac{dS}{dt} = -\beta(t) \frac{S(t)}{N - D(t)} I(t) \quad (3.13)$$

$$\frac{dI}{dt} = \beta(t) \frac{S(t)}{N - D(t)} I(t) - \gamma I(t) \quad (3.14)$$

$$\frac{dR}{dt} = (1 - \nu) \gamma I(t) \quad (3.15)$$

$$\frac{dD}{dt} = \nu \gamma I(t). \quad (3.16)$$

Equations (3.13)–(3.16) follow the progression of individuals in a population of size N between four different states: S , susceptible to the COVID-19 virus, I , infected with COVID-19 (both symptomatic and asymptomatic), R , recovered and no longer contagious, and D , deceased. In this model we assume that as agents move to category R , they acquire immunity to the disease for at least three months, which is the time frame considered in our experiments.

The parameter γ , called *recovery rate*, governs the evolution of infected people from state I to states R or D . Based on the severity of the disease, the value of γ is estimated to be somewhere between $1/4$ and $1/14$, which corresponds to the infectious period from 4 to 14 days [2]. In line with [2], we use $\gamma = 1/5$.

The second parameter, ν , is the fatality rate of the virus. Estimating the fatality rate is very difficult, since COVID-19 cases are believed to be substantially underreported (in part, due to a large number asymptomatic cases, especially among children and young adults). While early measurements from limited data for an emerging COVID-19 outbreak suggested ν to be as high as 1.2%, the more recent estimates based on antibody testing point toward a much lower value of 0.2% (though it does increase markedly with age and risk factors) [2]. Still, even with $\nu = 0.012$, the model grossly underestimates the actual number of deaths as shown by our numerical simulations below.

The *transmission rate*, $\beta(t)$, is defined as probability of infection given a contact between an infectious and susceptible individual multiplied by the average rate of contacts between these groups. It is the defining rate in disease progression and one of the two components in the *effective reproduction number*, $\mathcal{R}(t)$, the rate at which susceptible agents get infected divided by the recovery rate of infected individuals at time t , i.e.,

$$\mathcal{R}(t) = \frac{\beta(t)}{\gamma} \frac{S(t)}{N - D(t)}. \quad (3.17)$$

The effective reproduction number of a disease varies in time and it is directly affected by social response and public health guidelines, aimed at eventually bringing it under 1 for a sustained period of time needed to stop the chain of transmission. Below we demonstrate how iterative scheme (1.8) and (1.9) can be used to reconstruct these two important parameters, $\mathcal{R}(t)$ and $\beta(t)$. Previously, a time-dependent transmission rate, $\beta(t)$, has been approximated for various diseases from a similar (SEIR) compartmental model [24]. Numerous challenges faced by the authors while solving this parameter estimation problem motivated the development of new PCA (1.8) and (1.9), introduced in the current paper.

Let incidence data, d_δ , be reported on days $t_1, t_2, \dots, t_n$, where $\mathcal{A} < t_1 < t_2 < \dots < t_n < \mathcal{B}$. According to the model, the daily number of new incidence cases is equal to $\beta(t) \frac{S(t)I(t)}{N-D(t)} = \frac{dI}{dt} + \gamma I(t)$. Therefore one has to solve the following constrained minimization problem

$$\min_I \left\| \frac{dI}{dt} + \gamma I(t) - d_\delta \right\|^2 = \min_I \sum_{i=1}^n \left(\frac{dI}{dt}(t_i) + \gamma I(t_i) - d_\delta(t_i) \right)^2, \quad (3.18)$$

subject to (3.13)–(3.16) and initial conditions $S(t_1) = N - I^{(1)}$, $I(t_1) = I^{(1)}$, $R(t_1) = 0$, $D(t_1) = 0$. The goal of our experiment is to estimate the disease transmission rate, $\beta(t)$, by solving minimization problem (3.13)–(3.16), (3.18) with PCA (1.8) and (1.9) over the interval $[t_1, t_n]$, and then to use the reconstructed function $\beta(t)$ to calculate $\mathcal{R}(t)$ by formula (3.17).

To discretize $I'(t)$, we employ the same Fourier expansion as the one used in the previous subsection for discretizing $\frac{dC}{dt}$:

$$W(t) := I'(t) = \sum_{j=1}^N \left\{ A_j \left[\cos \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right] + B_j \sin \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) \right\}. \quad (3.19)$$

This discrete approximation will guarantee that $W(\mathcal{A}) = W(\mathcal{B}) = 0$. For $I(t)$, one obtains

$$\begin{aligned} I(t) = \int_{\mathcal{A}}^t W(s) ds &= \sum_{j=1}^N \left\{ A_j \left[\frac{\mathcal{B} - \mathcal{A}}{2\pi j} \sin \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - (t - \mathcal{A}) \right] \right. \\ &\quad \left. - B_j \frac{\mathcal{B} - \mathcal{A}}{2\pi j} \left[\cos \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right] \right\}. \end{aligned} \quad (3.20)$$

If one introduces the notation:

$$u := [A_1, \dots, A_N, B_1, \dots, B_N]^T, \quad (3.21)$$

then one can define $B(u)$ in (1.3) as a linear operator from $\mathbb{R}^{2N}$ to $\mathbb{R}^n$ (with n being the number of data points) such that

$$B_i(u) := W[u](t_i) + \gamma I[u](t_i), \quad i = 1, 2, \dots, n. \quad (3.22)$$

Furthermore, to recover the shape of $\beta(t)$, we project the transmission rate onto a finite subset spanned by the shifted Legendre polynomials of degree $0, 1, \dots, m-1$, which are orthogonal on the interval $[\mathcal{A}, \mathcal{B}]$ with respect to L_2 inner product, defined recursively as follows

$$\begin{aligned} x &= \frac{2t - \mathcal{A} - \mathcal{B}}{\mathcal{B} - \mathcal{A}}, \quad P_0(x) = 1, \quad P_1(x) = x, \quad t \in [\mathcal{A}, \mathcal{B}], \\ (j+1)P_{j+1}(x) &= (2j+1)xP_j(x) - jP_{j-1}(x), \quad j = 1, 2, \dots, m-2. \end{aligned}$$

This yields the following finite dimensional approximation of the transmission rate:

$$\beta(t) = \sum_{j=0}^{m-1} \theta_{j+1} P_j(t). \quad (3.23)$$

Representation (3.23) enables us to cast $G_i(\theta, u)$ in (1.3) as an operator from $\mathbb{R}^m \times \mathbb{R}^{2N} \rightarrow \mathbb{R}^n$:

$$G_i(\theta, u) := B_i(u) - \beta[\theta](t_i) \frac{N - I[u](t_i) - \gamma U[u](t_i)}{N - \nu \gamma U[u](t_i)} I[u](t_i) \quad \text{and} \quad g := 0, \quad (3.24)$$

with $U(t)$ defined in terms of $A_1, \dots, A_N, B_1, \dots, B_N$ as follows

$$U(t) := \int_{\mathcal{A}} I(s) ds = \sum_{j=1}^N \left\{ A_j \left[-\frac{(\mathcal{B} - \mathcal{A})^2}{(2\pi j)^2} \left\{ \cos \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - 1 \right\} - \frac{(t - \mathcal{A})^2}{2} \right] - B_j \frac{\mathcal{B} - \mathcal{A}}{2\pi j} \left[\frac{\mathcal{B} - \mathcal{A}}{2\pi j} \sin \left(2\pi j \frac{t - \mathcal{A}}{\mathcal{B} - \mathcal{A}} \right) - (t - \mathcal{A}) \right] \right\}. \quad (3.25)$$

Given (3.22) and (3.24), one can easily calculate Fréchet derivatives of $B(u)$ and $G(\theta, u)$ with respect to u and the derivative of $G(\theta, u)$ with respect to θ , and implement algorithm (1.8) and (1.9) for estimating $\beta(t)$ from (3.18) subject to (3.13)–(3.16). Considering the nature of θ , we do not scale any components in the penalty term of (1.8) and in all our simulations $T^*T = \mathcal{I}$, the identity operator in $\mathbb{R}^m$ [27].

We use a parametric bootstrap approach [7, 8, 11] to generate uncertainty bounds for reconstructed coordinates of θ , assuming Poisson error structure. We refit the model to each of the $M = 100$ additional data sets, generated by the bootstrap method, resulting in M best-fit parameter sets that are used to construct the 95% confidence intervals for each $\theta_j, j = 1, 2, \dots, m$. To ensure an unbiased choice of the initial guess for $\beta(t)$, we randomly select a constant θ_0 from the uniform distribution on $[0.1, 1]$, and take $[\theta_0, 0, 0, \dots, 0]^T$ to serve as initial approximation for the transmission rate expansion coefficients at every bootstrap iteration. Note that with this choice of θ , $\beta_0(t) = \theta_0$. To find an initial guess for u , we spline the approximation of $\frac{dI}{dt}(t_k)$:

$$\frac{dI}{dt}(t_k) \approx d_\delta(t_k) - \gamma \exp(-\gamma t_k) \sum_{j=1}^k \exp(\gamma t_j) d_\delta(t_j),$$

and compute its Fourier expansion coefficients $A_1^0, \dots, A_N^0, B_1^0, \dots, B_N^0$. We then set

$$u_0 := [A_1^0, \dots, A_N^0, B_1^0, \dots, B_N^0]^T.$$

For our numerical experiments, we use the same incidence data sets for COVID-19 virus in Netherlands and Spain as the sets used in the previous subsection. For Netherlands, in figure 8 and in table 1, the values of the reconstructed parameters, $\theta_j, j = 1, \dots, 10$, and the corresponding 95% confidence intervals (CIs) are presented. In figure 9 (right), a bundle of reconstructed values of the disease transmission rate, $\beta(t)$, can be seen along with randomly selected initial values for 100 bootstrap iterations. In figure 9 (left), the reconstructed incidence curves are illustrated. Figure 10 shows what our findings imply for the effective reproduction number, $\mathcal{R}(t)$. It starts off with a rather high level at the onset of the outbreak. Then about 47 days into the process (and shortly after the incidence curve peaks), the values of $\mathcal{R}(t)$ drop under 1, generating a declining flow of cases. The reproduction curve does, however, have a relapse toward

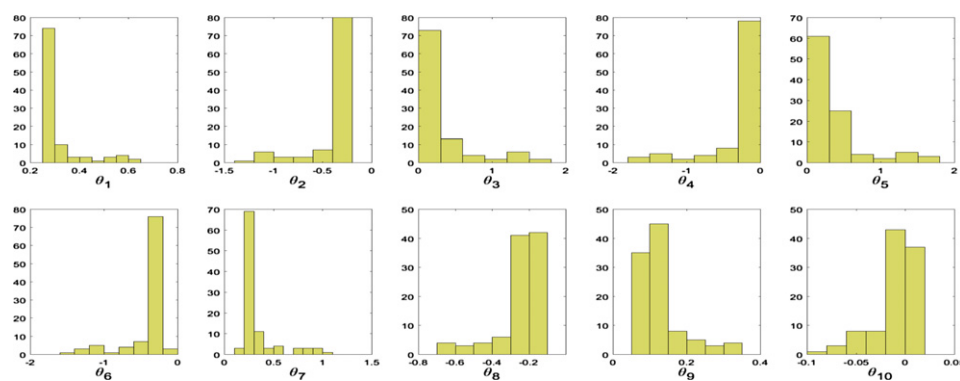


Figure 8. Netherlands: reconstructed values of Legendre coefficients for $\beta(t)$.

Table 1. Netherlands: 95% confidence intervals for the reconstructed Legendre coefficients.

$\theta_1 = 0.32(95\% \text{CI} : [0.28, 0.58])$	$\theta_2 = -0.38(95\% \text{CI} : [-1.1, -0.25])$
$\theta_3 = 0.4(95\% \text{CI} : [0.22, 1.4])$	$\theta_4 = -0.37(95\% \text{CI} : [-1.5, -0.15])$
$\theta_5 = 0.42(95\% \text{CI} : [0.19, 1.5])$	$\theta_6 = -0.4(95\% \text{CI} : [-1.3, -0.2])$
$\theta_7 = 0.34(95\% \text{CI} : [0.19, 0.96])$	$\theta_8 = -0.25(95\% \text{CI} : [-0.62, -0.15])$
$\theta_9 = 0.13(95\% \text{CI} : [0.079, 0.31])$	$\theta_{10} = -0.0098(95\% \text{CI} : [-0.067, 0.014])$

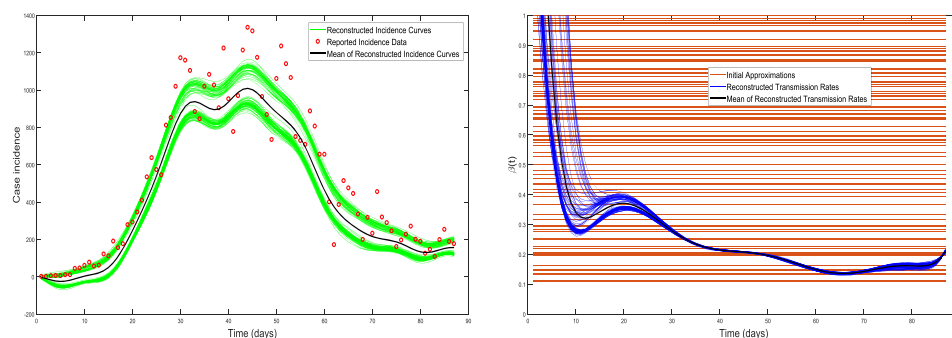


Figure 9. Netherlands: reconstructed incidence cases (left) and transmission rate (right).

the end of our study period, hinting that the second wave of the disease may be imminent. That prediction proved to be accurate [35].

In figure 11, the reconstructed functions $S(t)$, $I(t)$, $R(t)$, and $D(t)$ are given. The actual number of COVID-19-related deaths in Netherlands on May 24, 2020, was reported to be 5,811 [35], which is much higher than what the model projects. We face the same problem with $D(t)$ estimates in Spain (figure 12): the actual number of COVID-19-related deaths in Spain on June 7, 2020, was reported to be 28 323 [35], while the model projects slightly under 3,500. In the future, the model needs to be adjusted to include a percentage of reported cases as additional unknown parameter (due to a large number of unreported asymptomatic and mild cases) and/or to account for a time-dependent death rate whose magnitude is the highest at the early stages of the pandemic.

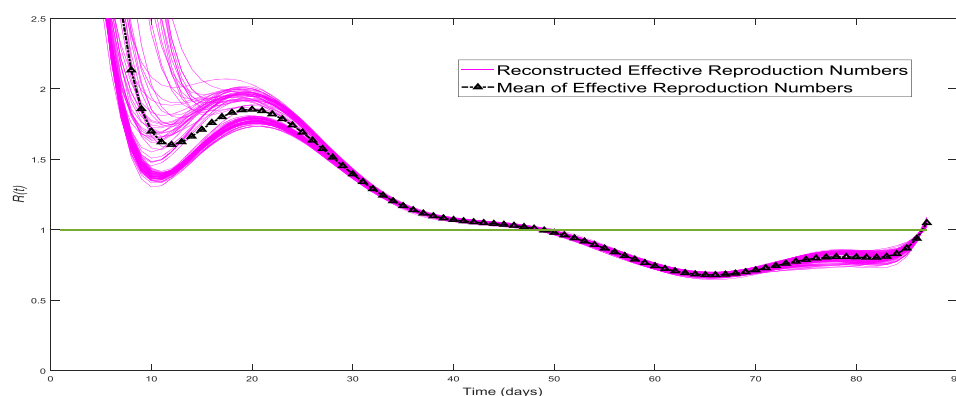


Figure 10. Netherlands: reconstructed values of effective reproduction number, $\mathcal{R}(t)$.

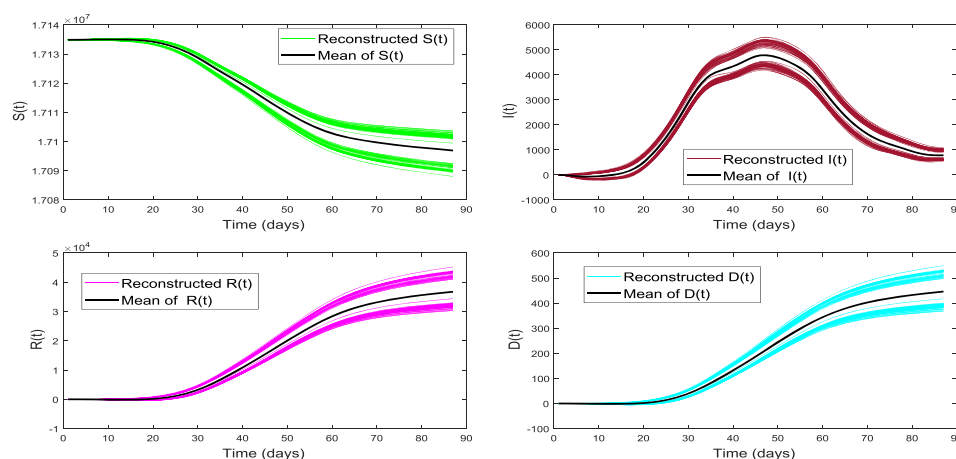
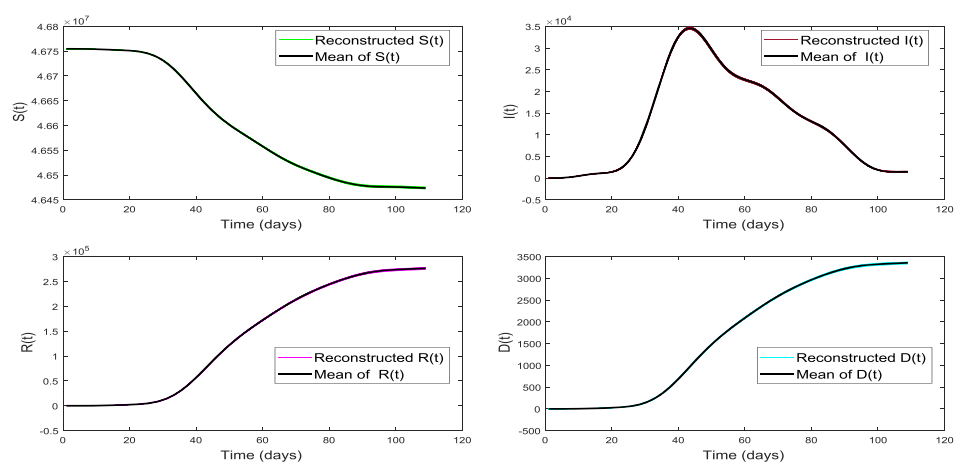
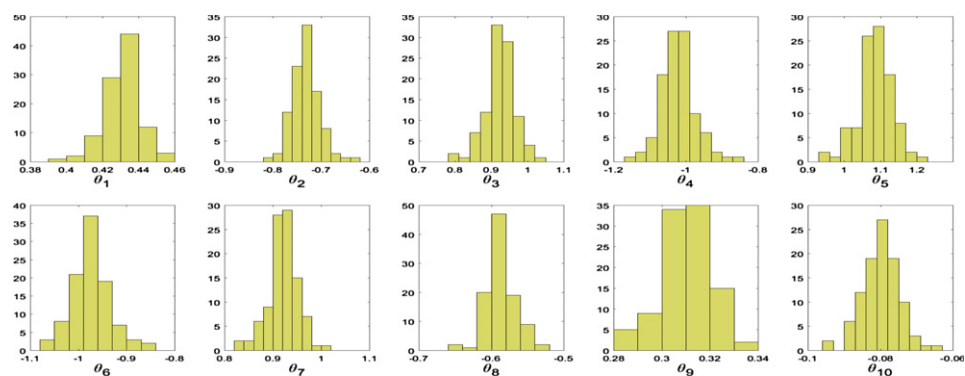


Figure 11. Netherlands: reconstructed values of $S(t)$, $I(t)$, $R(t)$, and $D(t)$.

In the case of Netherlands, for every bootstrap iteration, we take $\alpha_0 = 10^{-5}$, but a wide range of values from $\alpha_0 = 10^{-3}$ to $\alpha_0 = 10^{-8}$ can be used to get the results that are almost identical. The convergence rate for $\{\alpha_k\}$ is chosen to be $\alpha_k = \alpha_0/k$, the rate that gives rise to the most stable iterative process. Iterations are terminated once α_k has reached the value of 6.67×10^{-7} . We discretize $\frac{dI}{dt}(t)$ and $\beta(t)$ with 16 and 10 base functions, respectively (that is, $N = 8$ and $m = 10$).

Figure 13 and table 2 illustrate the values of the reconstructed parameters, θ_j , $j = 1, \dots, 10$, and the corresponding 95% confidence intervals (CIs), given COVID-19 incidence data for Spain. In figure 14 (right), a bundle of reconstructed values of the disease transmission rate, $\beta(t)$, is shown along with randomly selected initial values for 100 bootstrap iterations. In figure 14 (left), the reconstructed incidence curves are presented. In figure 15, the effective reproduction number, $\mathcal{R}(t)$, for the Spain data is demonstrated. Like in the case of Netherlands, it starts off with a rather high level at the onset of the outbreak. Between days 25 and 100, the effective reproduction number for Spain decreases from 2 to 0.5. Then it begins to

Figure 12. Spain: reconstructed values of $S(t)$, $I(t)$, $R(t)$, and $D(t)$.Figure 13. Spain: reconstructed values of Legendre coefficients for $\beta(t)$.**Table 2.** Spain: 95% confidence intervals for the reconstructed Legendre coefficients.

$\theta_1 = 0.43(95\%CI : [0.41, 0.45])$	$\theta_2 = -0.73(95\%CI : [-0.79, -0.67])$
$\theta_3 = 0.92(95\%CI : [0.84, 1])$	$\theta_4 = -1(95\%CI : [-1.1, -0.92])$
$\theta_5 = 1.1(95\%CI : [0.99, 1.2])$	$\theta_6 = -0.97(95\%CI : [-1, -0.89])$
$\theta_7 = 0.92(95\%CI : [0.86, 0.98])$	$\theta_8 = -0.59(95\%CI : [-0.62, -0.54])$
$\theta_9 = 0.31(95\%CI : [0.29, 0.33])$	$\theta_{10} = -0.08(95\%CI : [-0.089, -0.071])$

grow and reaches the level of 1.4 toward the end of the study period suggesting a possibility of the second wave. Again, this prognosis proved to be accurate [35].

For Spain incidence data, we use $\alpha_0 = 10^{-10}$ at every bootstrap iteration (a range of values from 10^{-5} to 10^{-12} results in very similar solutions). The convergence rate for $\{\alpha_k\}$ is chosen to be $\alpha_k = \alpha_0/k$. The process is terminated once α_k has reached the value of 2×10^{-11} , i.e., after 5 iterations. We discretize both $\frac{dI}{dt}(t)$ and $\beta(t)$ with 10 base functions (that is, $N = 5$ and $m = 10$). This rather aggressive discretization of the state variable allows to reduce the value of α_0 as compared to our previous experiment, where $N = 8$ and $\frac{dI}{dt}(t)$ is discretized with 16 base

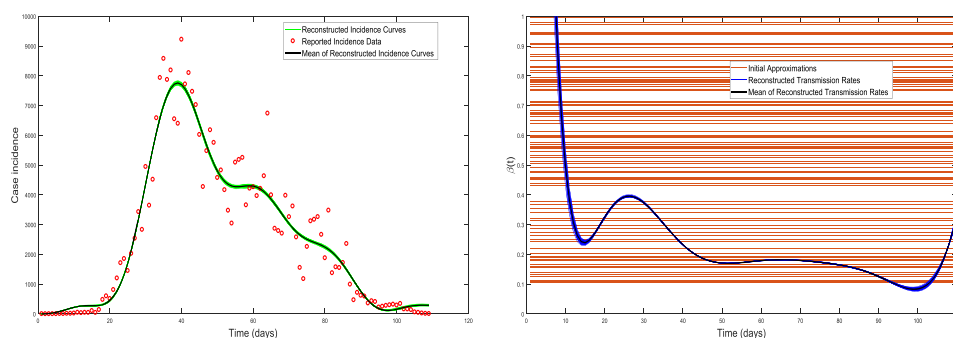


Figure 14. Spain: reconstructed incidence cases (left) and transmission rate (right).

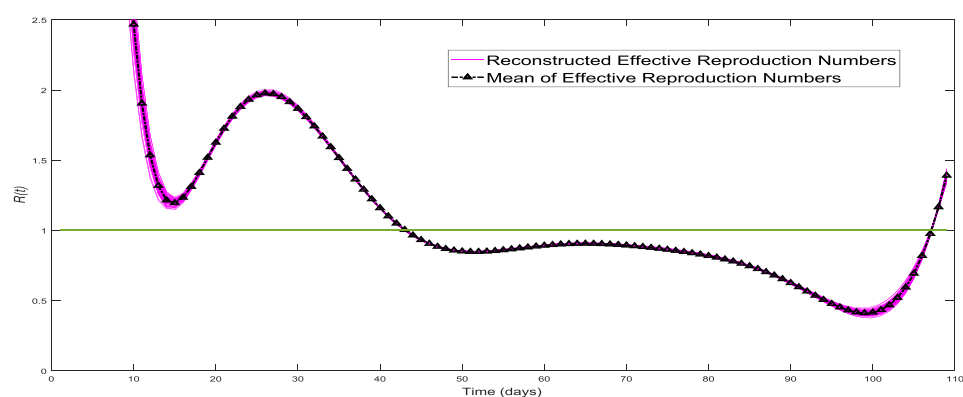


Figure 15. Spain: reconstructed values of effective reproduction number, $\mathcal{R}(t)$.

functions. It also allows to reduce the number of iterations from 15 to 5, and to recover $\beta(t)$ and $\mathcal{R}(t)$ with very low uncertainty (see figures 14 and 15). On the other hand, if one uses $N \geq 8$ with Spain incidence data, then, regardless of the value of α_0 , uncertainty in the recovered $\beta(t)$ tends to be rather high near the end points.

Numerical simulations presented in this subsection demonstrate another important advantage of our proposed algorithm (1.8) and (1.9). Indeed, if one solves (3.18) subject to (3.13)–(3.16) in a traditional way [24], then one has to evaluate the operator $J(\theta) := B(u(\theta))$, which is a composition of the parameter-to-state map, $u = u(\theta)$, satisfying $G(\theta, u(\theta)) = g$, and the observation operator, $B = B(u)$. That is, $J(\theta) := B(u(\theta))$. As it is clear from (3.13)–(3.16), u is a nonlinear function of θ , which implies that $J(\theta)$ is also nonlinear. On the other hand, our PCA does not require solving $G(\theta, u) = g$ for u , and the operator $G(\theta, u)$ is linear with respect to θ (though nonlinear with respect to u), which results in much simpler iterations for θ_k as compared to the traditional approach (without any increase in the solution space for this unknown parameter).

4. Conclusions and discussion

New PCA (1.8) and (1.9) has been theoretically justified for nonlinear minimization problem (1.3). Numerical simulations have been carried out for two inverse problems in epidemiology, aimed at parameter estimation and forecasting of future incidence cases. The experiments

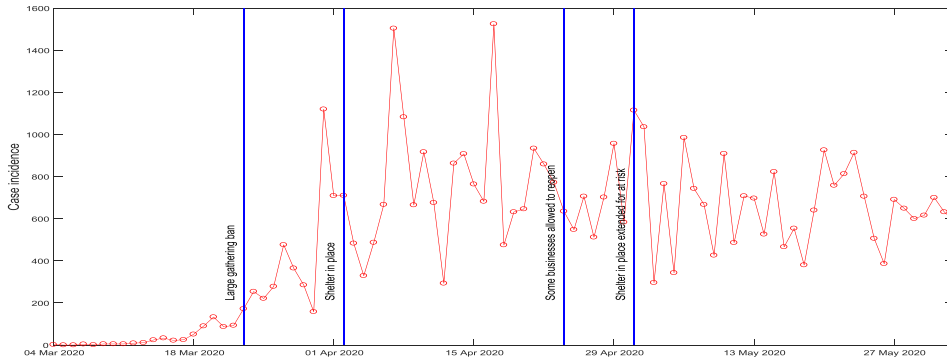


Figure 16. The state of Georgia, USA, COVID-19 incidence data.

have shown that our new method is accurate and stable for a broad range of initial values and regularization sequences. As compared to the ‘traditional’ approach [25], which consists in minimizing $\|B(u) - d_\delta\|^2$ with respect to θ , while solving the equation $G(\theta, u) = g$ numerically at every step of the iterative process, for the two inverse problems considered, method (1.8) and (1.9) is more time efficient and more reliable in its estimation of the unknown parameters.

From theoretical standpoint, one of the main assumptions of the convergence theorems is condition (1.3) on the well-posedness of the forward problem. Our numerical experiments confirm that this assumption is reasonable and that iterations (1.9), indeed, do not require any regularization except for the discrete approximation of a state variable.

An important topic that needs to be studied next is the extension of the modified Richards model to the case of elaborate epidemic trajectories aggregating multiple asynchronous sub-epidemics, since a large number of COVID-19 incidence data curves do not have a simple bell-shape behavior (see COVID-19 data for the state of Georgia, USA, in figure 16, for example [6]). In [7, 23], a complex outbreak structure comprised of $\mathcal{M}$ overlapping sub-epidemics is modeled as

$$\frac{dC_i}{dt} = rA_i(t)C_i^p(t) \left[1 - \frac{C_i(t)}{K_i} \right], \quad (6.1)$$

where $C_i(t)$ is the cumulative number of infections in sub-epidemic i ($i = 1, \dots, \mathcal{M}$), K_i is the size of the i th sub-epidemic, and the growth rate r and the scaling parameter p are the same across all $\mathcal{M}$ sub-waves [7]. The timing of onset for each consecutive sub-epidemic is modeled in such a way that the $(i+1)$ th sub-epidemic is triggered when the cumulative case count of sub-epidemic i , $C_i(t)$, exceeds a certain threshold, C_{thr} . The $(i+1)$ th sub-epidemic begins before the i th sub-epidemic reaches its extinction. The size of sub-epidemic i is modeled as $K_i = K_0 e^{-q(i-1)}$.

It is our expectation that model (6.1) could help to overcome some shortcomings of MR equation (3.1) and, combined with algorithm (1.8) and (1.9), could give rise to a robust forecasting tool that will be used to assess future viral infections.

As a long term goal, it would be of great interest to consider a machine learning (ML) formulation of the optimization problem [34], where the explicit form of $G(\theta, u)$ is unknown and one would like to ‘learn’ it. In this framework, one models $G(\theta, u)$ as a deep neural network, and θ consists of network parameters from multiple layers. ‘Training’ neural network, i.e., minimizing the functional $\sum_{i=1}^N \|B(u(g_i; \theta)) - d_i\|^2$ with respect to θ , subject to $G(\theta, u) = g_i$,

is very data demanding, since one needs a set of initial values $\{g_i\}$ and the corresponding trajectories $\{u(g_i; \theta)\}$. Having multiple g_i can be difficult in epidemiology, in general. However, with the abundance of COVID-19 data and with the virus being so widespread, one can take g_i as the initial condition of the i th outbreak in a particular country (or state) with d_i being incidence data for this country (state). ‘Learning’ the full model would be of huge advantage, since all compartmental and phenomenological models are simplifications, which are, in fact, also unknown to some degree.

Acknowledgment

The authors would like to express their sincere gratitude to the anonymous referees for taking the time to read this manuscript very carefully and for making brilliant comments and recommendations that helped to considerably improve the quality of this paper.

ORCID iDs

Alexandra Smirnova  <https://orcid.org/0000-0002-3135-304X>

References

- [1] Aster R C, Borchers B and Thurber C H 2011 *Parameter Estimation and Inverse Problems* (New York: Academic)
- [2] Atkeson A, Kopecky K and Zha T 2020 Estimating and forecasting disease scenarios for COVID-19 with an SIR model *NBER Working Paper No. 27335, Issued in June 2020 NBER Program(s): Economic Fluctuations and Growth*
- [3] Bakushinsky A B 1993 Iterative methods for nonlinear operator equations without regularity. New approach *Dokl. Russian Acad. Sci.* **330** 282–4
- [4] Bakushinsky A B and Kokurin M Y 2004 *Iterative Methods for Ill-Posed Operator Equations with Smooth Operators* (Berlin: Springer)
- [5] Bakushinsky A B and Leonov A S 2020 Numerical solution of an inverse multifrequency problem in scalar acoustics *Comput. Math. Math. Phys.* **60** 987–99
- [6] CDC COVID Data Tracker <https://covid.cdc.gov/covid-data-tracker/#cases>
- [7] Chowell G, Tariq A and Hyman J M 2019 A novel sub-epidemic modeling framework for short-term forecasting epidemic waves *BMC Med.* **17** 164
- [8] Chowell G 2019 Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts *Infect. Dis. Model.* **2** 379–98
- [9] Chowell G, Viboud C, Hyman J M and Simonsen L 2015 The Western Africa Ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates *PLoS Curr* **21** 7
- [10] Chowell G, Ammon C E, Hengartner N W and Hyman J M 2006 Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: assessing the effects of hypothetical interventions *J. Theor. Biol.* **241** 193–204
- [11] Efron B and Tibshirani R 1986 Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy *Stat. Sci.* **1** 54–75
- [12] Giordano G, Blanchini F, Bruno R, Colaneri P, Colaneri P, Di Filippo A, Di Matteo A and Colaneri M 2020 Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy *Nat. Med.* **26** 855–60
- [13] Hanke M 1997 Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems *Numer. Funct. Anal. Optim.* **18** 971–93
- [14] Engl H, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
- [15] Kaltenbacher B 2018 Minimization based formulations of inverse problems and their regularization *SIAM J. Optim.* **28** 620–45
- [16] Kaltenbacher B 2016 Regularization based on all-at-once formulations for inverse problems *SIAM J. Numer. Anal.* **54** 2594–618

- [17] Kaltenbacher B, Kirchner A and Vexler B 2014 Goal oriented adaptivity in the IRGNM for parameter identification in PDEs: II. all-at-once formulations *Inverse Problems* **30** 045002
- [18] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems (Radon Series on Computational and Applied Mathematics vol 6)* (Berlin: de Gruyter)
- [19] Nocedal J and Wright S 2000 *Numerical Optimization* (Berlin: Springer)
- [20] Neubauer A 2020 Optimal convergence rates for inexact Newton regularization with CG as inner iteration *J. Inverse Ill-Posed Probl.* **28** 145–53
- [21] Ortega J M and Rheinboldt W C 2014 *Iterative Solution of Nonlinear Equations in Several Variables* (New York: Academic)
- [22] Jin Q and Wang W 2018 Analysis of the iteratively regularized Gauss-Newton method under a heuristic rule *Inverse Problems* **34** 035001
- [23] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J M, Yan P and Chowell G 2020 Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 1323, 2020 *J. Clin. Med.* **9** 596
- [24] Smirnova A, DeCamp L and Chowell G 2019 Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model *Bull. Math. Biol.* First Online: 02 May 2017, ISSN (Online) 1522-9602, ISSN (Print) 0092-8240 **81** 4343–65
- [25] Smirnova A, Chowell-Puente G, DeCamp L, Moghadas S and Sheppard M J 2017 Improving epidemic size prediction through stable reconstruction of disease parameters by reduced iteratively regularized Gauss–Newton algorithm *J. Inverse Ill-Posed Probl.* **25** 653–67
- [26] Smirnova A 2012 On convergence rates for iteratively regularized procedures with linear penalty terms *Inverse Problems* **28** 085005
- [27] Smirnova A, Renaut R A and Khan T 2007 Convergence and application of a modified iteratively regularized Gauss–Newton algorithm *Inverse Problems* **23** 1547–63
- [28] Thompson R N 2020 Epidemiological models are important tools for guiding COVID-19 interventions *BMC Med.* **18** 152
- [29] Tikhonov A N, Goncharsky A, Stepanov V V and Yagola A G 1995 *Numerical Methods for the Solution of Ill-Posed Problems (Mathematics and its Applications vol 328)* (Berlin: Springer) p 251
- [30] Tsoularis A and Wallace J 2001 Analysis of logistic growth models *Math. Biosci.* **179** 21–55
- [31] Turner M E, Bradley E L, Kirk K A and Pruitt K M 1976 A theory of growth *Math. Biosci.* **29** 367–73
- [32] Vasin V V and Ageev A L 1995 *Ill-posed Problems with a priori Information* (Utrecht: VNU)
- [33] Werner F and Hofmann B 2020 Convergence analysis of (statistical) inverse problems under conditional stability estimates *Inverse Problems* **36** 23
- [34] Wiemken T L and Kelley R R 2020 Machine learning in epidemiology and health outcomes research *Annu. Rev. Public Health* **41** 21–36
- [35] COVID-19 Situation in the WHO European Region data as of: 12 October 2020, 11:18am CEST <https://covid19.who.int/region/euro/country/es>