On inertial Levenberg-Marquardt type methods for solving nonlinear ill-posed operator equations

A. Leitão[†] D.A. Lorenz[‡] J.C. Rabelo[§] M. Winkler[¶]

June 8, 2024

Abstract

In these notes we propose and analyze an inertial type method for obtaining stable approximate solutions to nonlinear ill-posed operator equations. The method is based on the Levenberg-Marquardt (LM) iteration. The main obtained results are: monotonicity and convergence for exact data, stability and semi-convergence for noisy data. Regarding numerical experiments we consider: i) a parameter identification problem in elliptic PDEs, ii) a parameter identification problem in machine learning; the computational efficiency of the proposed method is compared with canonical implementations of the LM method.

Keywords. Ill-posed problems; Nonlinear equations; Two-point methods; Inertial methods; Levenberg Marquardt method.

AMS Classification: 65J20, 47J06.

1 Introduction

sec:intro

In a standard inverse problem scenario [4, 10, 17], consider Hilbert spaces X and Y and contemplate the challenge of deducing an unknown quantity $x \in X$ from provided data $y \in Y$. In other words, the task is to identify an unknown quantity of interest x (which cannot be directly accessed) relying on information derived from a set of measured data y.

An essential aspect to note is that, in real-world applications, the precise data $y \in Y$ is not accessible. Instead, only approximate measured data $y^\delta \in Y$ is at our disposal, meeting the criteria of

$$\|y^\delta - y\| \ \leq \ \delta \, . \tag{1}$$

eq:noisy-

Here, $\delta > 0$ represents the level of noise and we assume that δ (or an estimate thereof) is known. The available noisy data $y^\delta \in Y$ are obtained by indirect measurements of $x \in X$, this process being represented by the model

$$F(x) \ = \ y^\delta , \tag{2}$$

eq:inv-pr

where $F : X \to Y$, is a nonlinear, Fréchet differentiable, ill-posed operator.

[†]Department of Mathematics, Federal Univ. of St. Catarina, 88040-900 Floripa, Brazil
[‡]Center for Industrial Mathematics, University of Bremen, Bibliotheksstrasse 5, 28359 Bremen, Germany
[§]Department of Mathematics, Federal Univ. of Piaui, 64049-550 Teresina, Brazil
[¶]Center for Industrial Mathematics, University of Bremen, Bibliotheksstrasse 5, 28359 Bremen, Germany
Emails: acgleitao@gmail.com, d.lorenz@uni-bremen.de, joelrabelo@ufpi.edu.br, maxwin@uni-bremen.de.

1.1 State of the art

The Levenberg-Marquardt (LM) method We recall a family of implicit iterative type methods for obtaining stable approximate solutions to nonlinear ill-posed type operator equations as in (2). The Levenberg-Marquardt (LM) type methods are defined by

def:LM

$$x_{k+1}^\delta := \text{argmin}_x \left\{ \|F(x_k^\delta) + F'(x_k^\delta)(x - x_k^\delta) - y^\delta\|^2 + \lambda_k \|x - x_k^\delta\|^2 \right\}, \ k = 0, 1, \dots, \quad \text{(3a)}$$

what corresponds to defining x_{k+1}^δ as the solution of the optimality condition

$$\left(A_k^* A_k + \lambda_k I\right)(x - x_k^\delta) \ = \ A_k^* \left(y^\delta - F(x_k^\delta)\right), \ k = 0, 1, \dots \quad \text{(3b)}$$

where $A_k := F'(x_k^\delta) : X \to Y$ is the Fréchet derivative of F evaluated at x_k^δ and $A_k^* : Y \to X$ is the adjoint operator to A_k. Additionally, (λ_k) is a positive sequence of Lagrange multipliers. The iteration starts at a given initial guess $x_0 \in X$.

In the case of linear ill-posed operator equations (i.e. $F(x) = Ax$; notice that (2) becomes $Ax = y^\delta$) the LM method reduces to the iterated Tikhonov method [13, 6] (or proximal point method [19, 27]), which correspond to defining $x_{k+1}^\delta := \arg\min_x \left\{ \|Ax - y^\delta\|^2 + \lambda_k \|x - x_k^\delta\|^2 \right\}$. The parameters λ_k are appropriately chosen Lagrange multipliers [6].

The literature on LM type methods for inverse problems is extensive, exploring various aspects, including regularization properties [9, 12, 16, 15], convergence rates [13, 28], *a posteriori* strategies for choosing the Lagrange multipliers [6], a cyclic version of the LM method [8], among others.

Inertial iterative methods Inertial iterative methods have been introduced by Polyak in [23] for the minimization of a smooth convex function f. The algorithm is written as a two step method

$$\begin{aligned} w_k &= x_k + \alpha_k (x_k - x_{k-1}) \\ x_{k+1} &= w_k - \lambda_k \nabla f(x_k) \end{aligned}$$

where α_k is an extrapolation between 0 and 1 and λ_k is a stepsize. The method is called the heavy-ball method as the extrapolation can be motivated by a discretization of the dynamical system $\ddot{x}(t) + \gamma \dot{x}(t) = -\nabla f(x(t))$ which models the dynamics of a mass with friction driven by a potential f. The method has also been extended to monotone operators, e.g. by Alvarez and Attouch in [1] for the proximal point method and by Moudafi and Oliny in [20] for the forward-backward method.

The heavy ball method achieves the optimal lower complexity bounds for first order methods for smooth strongly convex functions [22]. For merely smooth function, a simple modification proposed by Nesterov in [21] achieves the lower complexity bounds also in this case. The method reads as

$$\begin{aligned} w_k &= x_k + \alpha_k (x_k - x_{k-1}) \\ x_{k+1} &= w_k - \lambda_k \nabla f(w_k) \end{aligned}$$

and the only difference to the heavy ball method is that the gradient is also evaluated at the extrapolated point. The performance relies on a clever choice of the extrapolation sequence α_k such that it approaches 1 not too fast and not too slow. The method has been extended to the forward backward case for convex optimization by Beck and Teboulle [5] and further to monotone inclusions by Lorenz and Pock [18]. Su, Boyd and Candés [29] related Nesterov's method to the dynamical system $\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = -\nabla f(x(t))$ which is is similar to the heavy ball method but the damping α/t vanishes asymptotically. The viewpoint of continuous dynamics was further elaborated by Alvarez, Attouch, Bolte and Redont [2] where the authors proposed to analyze

$$\ddot{x}(t) + (\alpha I + \beta\nabla^2 f(x(t)))\ddot{x}(t) = -\nabla f(x(t))$$

which they called *dynamic inertial Newton system* (DIN). After time disretization, this leads to an inertial Levenberg-Marquardt method similar to the one we consider in this paper, but [2] only analyzed the continuous time system. Attouch, Peypoquet and Redont [3] combined the DIN method with vanishing damping

$$\ddot{x}(t) + (\tfrac{\alpha}{t} I + \beta\nabla^2 f(x(t)))\ddot{x}(t) = -\nabla f(x(t)).$$

1.2 Contribution

In these notes, we introduce and analyze an implicit inertial iteration, here called *inertial Levenberg-Marquardt method* (inLM), which can be construed as an extension of the LM method. Our approach is connected to the inertial method put forth in 2001 by Alvarez and Attouch [1]. In the case of linear ill-posed operator equations an approch analog to the one addressed in this manuscript (namely the *inertial iterated Tikhonov method*) is treated in [24].

We suggest this implicit inertial method as a practical alternative for computing robust approximate solutions to the ill-posed operator equation (2) and explore its numerical effectiveness.

The method under consideration consists in choosing appropriate non-negative sequences (α_k), (λ_k) and defining (at each iterative step) the extrapolation $w_k^\delta := x_k^\delta + \alpha_k(x_k^\delta - x_{k-1}^\delta)$; the next iterate x_{k+1} is than defined by

$$x_{k+1}^\delta := \operatorname{argmin}_x \left\{ \|F(w_k^\delta) + F'(w_k^\delta)(x - w_k^\delta) - y^\delta\|^2 + \lambda_k \|x - w_k^\delta\|^2 \right\}, \ k = 0, 1, \dots \qquad (4)$$ eq:inLM-s

where $x_{-1} = x_0 \in X$ are given. For obvious reasons we refer to this implicit two-point method as *inertial Levenberg-Marquardt method* (inLM).

1.3 Outline

The outline of the manuscript is as follows: In Section 2 we introduce and analyze the inLM method. We prove a monotonicity result as well as convergence for exact data in Section 2.2, and discuss stability and semi-convergence results in Section 2.3. In Section 3 the inLM method is tested for two ill-posed problems: i) a parameter identification problem in elliptic PDEs, ii) a parameter identification problem in machine learning. Section 4 is devoted to final remarks and conclusions.

2 The inertial Levenberg-Marquardt method

sec:iteration

In this section we introduce and analyze the inLM method considered in these notes. In Section 2.1 the inLM method is presented and preliminary results are derived. A convergence result (in the exact data case) is proven in Section 2.2. Stability and semi-convergence results (in the noisy data case) are proven in Section 2.3.

This is the set of main assumptions that we impose on the operator F and the data y:

(A1) The operator $F : X \to Y$ is continuously Fréchet differentiable. Moreover, there exist constants $C > 0$ and $\rho > 0$ such that $\|F'(x)\| \le C$, for all $x \in B_\rho(x_0)$.

(A2) The operator F satisfies the weak Tangential Cone Condition (wTCC) at $B_\rho(x_0)$ for some $\eta \in [0,1)$, i.e.

$$\|F(x') - F(x) - F'(x)(x'-x)\| \;\le\; \eta\,\|F(x') - F(x)\|\,, \quad \forall\, x,\, x' \in B_\rho(x_0).$$

(A3) There exists $x^\star \in B_{\rho/2}(x_0)$ such that $F(x^\star) = y$, where $y \in Rg(F)$ is the exact data.

(A4) There exists $q \in (\eta, 1)$ such that $\lambda_k > q\,C^2\,(1-q)^{-1}$, for $k = 0, 1, \ldots$

ssec:2.1

2.1 Description of method

To emphasize the fundamental principles that underlie the definition of our method, we commence the discussion by examining the scenario with exact data $y^\delta = y$, i.e. $\delta = 0$. Denoting the current iterate by $x_k \in X$, for $k \ge 0$, the step of the proposed inLM method consists in two parts: (i) compute $w_k \in X$, according to def:inLM

$$w_k \;:=\; x_k + \alpha_k\,(x_k - x_{k-1})\,; \tag{5a}$$ def:inLM-

(ii) define the subsequent iterate $x_{k+1} \in X$ as the solution of

$$\big(A_k^* A_k + \lambda_k I\big)\,(x - w_k) \;=\; A_k^*\,\big(y - F(w_k)\big)\,, \tag{5b}$$ def:inLM-

for $k = 0, 1, \ldots,$ were $A_k := F'(w_k) : X \to Y$ is the Fréchet derivative of F at w_k and $A_k^* : Y \to X$ is the adjoint operator to A_k. Here $x_0 \in X$ plays the role of an initial guess and $x_{-1} := x_0$. Moreover, $(\alpha_k) \in [0, \alpha)$ for some $\alpha \in (0,1)$, and $(\lambda_k) \in \mathbb{R}^+$ are given sequences. Notice that, if $\alpha_k \equiv 0$ then $w_k = x_k$ in (5a); thus, (5b) reduces to the standard LM iteration for exact data, i.e. x_{k+1} is defined as the solution of $\big(A_k^* A_k + \lambda_k I\big)(x - x_k) \;=\; A_k^*\,\big(y - F(x_k)\big)$, for $k = 0, 1, \ldots$.

The careful reader observes that (5b) is essentially the LM iterative step (3) starting from the extrapolation point w_k instead of x_k. Notice that (5b) is equivalent to computing

$$s_k \;:=\; \big(A_k^* A_k + \lambda_k I\big)^{-1} A_k^*\,\big(y - F(w_k)\big) \quad \text{and setting} \quad x_{k+1} := w_k + s_k \tag{6}$$ def:sk

(s_k is the iterative step of the inLM method). It is straightforward to see that the first equation in (6) is equivalent to $G_k\, s_k = A_k^*(y - F(w_k))$, where $G_k := (A_k^* A_k + \lambda_k I) : X \to X$ is a positive definite operator with spectrum contained in the interval $[\lambda_k, \lambda_k + \|A\|^2]$. Consequently, since $\lambda_k > 0$, the iterate x_{k+1} is uniquely defined by (5b).

We present the inLM method in algorithmic form in Algorithm 1.

rem:station **Remark 2.1** (Comments on Algorithm 1)**.** *This algorithm generates infinite sequences $(x_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$ if and only if $F(w_k) \ne y$, for all $k \in \mathbb{N}$. Indeed, if $F(w_{k_0}) = y$ for some $k_0 \in \mathbb{N}$ in Algorithm 1, the iteration stops at Step [2.4] after computing $x_0, \ldots, x_{k_0+1}$ and $w_0, \ldots, w_{k_0}$.*

The operators $A_k := F'(w_k) \in \mathcal{L}(X,Y)$ and $A_k^* \in \mathcal{L}(Y,X)$ do not have to be explicitely known (see the inverse problem in Section 3.2). The linear system in Step [2.1] can be solved, e.g., using the Conjugate Gradient (CG) mehthod; in this case it is enough to know only the action of A_k and A_k^*.

In the remaining of this subsection we establish preliminary properties of the sequences (x_k), (w_k) generated by Algorithm 1. The first result, stated in Lemma 2.2, follows directly from the definition of w_k in (5a) (see also Step [2.4] of Algorithm 1), while in Lemma 2.3 some useful inequalities are derived.

lemma:aux **Lemma 2.2.** *Let (A1) hold and (x_k), (w_k) be sequences generated by Algorithm 1. Thus*

$$\|w_k - x\|^2 \;=\; (1+\alpha_k)\|x_k - x\|^2 - \alpha_k\|x_{k-1} - x\|^2 + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2, \;\; k \ge 1 \tag{7}$$ eq:w

for $x \in X$.

[0] choose an initial guess $x_0 \in X$; $w_0 = x_0$; $k := 0$;

[1] choose $\alpha \in [0,1)$ and $(\lambda_k)_{k\geq 0} \in \mathbb{R}^+$;

[2] **for** $k = 0, 1, \dots$ **do**

if $\big(\|F(w_k) - y\| > 0\big)$ **then**

[2.1] $A_k := F'(w_k)$; compute $s_k \in X$ as the solution of

$$(A_k^* A_k + \lambda_k I)\, s_k \;=\; A_k^* \big(y - F(w_k)\big);$$

[2.2] $x_{k+1} := w_k + s_k$;

[2.3] choose $\alpha_{k+1} \in [0, \alpha]$; $w_{k+1} := x_{k+1} + \alpha_{k+1}(x_{k+1} - x_k)$;

else

[2.4] $s_k := 0$; $x_{k+1} := w_k$; **break**;

end if;

end for;

alg:init-exact

Algorithm 1: Inertial Levenberg-Marquardt method in the exact data case.

Proof. See [24, Lemma 2.2] for a complete proof. □

lemma:dk-and-ineq

Lemma 2.3. *Let (A1) hold and (x_k), (w_k) be sequences generated by Algorithm 1. Define $A_k := F'(w_k)$ and $D_k := F(w_k) + A_k(x_{k+1} - w_k) - y$. The following assertions hold true:*

a) $D_k \;=\; \lambda_k (A_k A_k^* + \lambda_k I)^{-1} (F(w_k) - y)$*;*

b) $w_k - x_{k+1} \;=\; \lambda_k^{-1} A_k^* \left[F(w_k) + A_k(x_{k+1} - w_k) - y\right]$*;*

c) Additionally, if (A4) holds, we have $q\|F(w_k) - y\| \;\leq\; \|D_k\| \;\leq\; \|F(w_k) - y\|$*;*

d) Additionally, if (A2) holds and w_k*,* $x_{k+1} \in B_\rho(x_0)$*, we have*

$$(1 - \eta)\|F(x_{k+1}) - y\| \;\leq\; (1 + \eta)\|F(w_k) - y\|.$$

Proof. **Assertions (a) and (b):** From Steps [2.1] and [2.2] of Algorithm 1 follow

$$A_k^* \left[F(w_k) + A_k(x_{k+1} - w_k) - y\right] + \lambda_k (x_{k+1} - w_k) \;=\; 0 \tag{8}$$ eq:stepLM

(see also (4)). Consequently, $A_k A_k^* D_k + \lambda_k A_k (x_{k+1} - w_k) = 0$, from where we obtain

$$A_k A_k^* D_k + \lambda_k A_k (x_{k+1} - w_k) + \lambda_k (F(w_k) - y) - \lambda_k (F(w_k) - y) \;=\; 0\,.$$

Thus, $(A_k A_k^* + \lambda_k I)\, D_k = \lambda_k (F(w_k) - y)$ and Assertion (a) follows.

Assertion (b) is an immediate consequence of (8).

Assertion (c): If (A1) and (A4) hold, we conclude from Assertion (a) together with the fact $\sigma(A_k A_k^* + \lambda_k I) \subset [\lambda_k, \lambda_k + C^2]$ that

$$\frac{\lambda_k}{\lambda_k + C^2} \|F(w_k) - y\| \;\leq\; \|D_k\| \;=\; \|\lambda_k (A_k A_k^* + \lambda_k I)^{-1} (F(w_k) - y)\| \;\leq\; \|F(w_k) - y\|\,.$$

From this inequality Assertion (c) follows.

Assertion (d): From the definition of D_k follows

$$\|F(x_{k+1}) - y\| \;\leq\; \|F(x_{k+1}) - F(w_k) - A_k(x_{k+1} - w_k)\| + \|D_k\|\,.$$

Thus, it follows from (A2) and Assertion (c)

$$\begin{aligned} \|F(x_{k+1}) - y\| \;&\leq\; \eta\, \|F(x_{k+1}) - F(w_k)\| + \|F(w_k) - y\| \\ &\leq\; \eta\, \|F(x_{k+1}) - y\| + (1 + \eta)\|F(w_k) - y\|), \end{aligned}$$

proving Assertion (d). □

ass:alpha **Assumption 2.4.** *Given $\alpha \in [0,1)$ and a convergent series $\sum_k \theta_k$ of nonnegative terms, let*

$$\alpha_k := \begin{cases} \min\left\{ \frac{\theta_k}{\|x_k - x_{k-1}\|^2}, \frac{\min\{\theta_k,\ \rho - \|x_k - x_0\|\}}{\|x_k - x_{k-1}\|}, \alpha \right\} & , \text{ if } \|x_k - x_{k-1}\| > 0 \\ 0 & , \text{ otherwise} \end{cases} , \ k \geq 1\,.$$

For simplicity of the presentation we assume, for the rest of this section, that $\theta_k = 1/k^2$.

rem:alpha **Remark 2.5.** *Should the sequence (α_k) of inertial parameters be chosen in accordance with Assumption 2.4, two immediate consequences ensue, namely:*

a) If x_{k-1}, $x_k \in B_\rho(x_0)$, then $w_k \in B_\rho(x_0)$ as well. Indeed, from (5a) follows

$$\|w_k - x_0\| \leq \|x_k - x_0\| + \alpha_k \|x_k - x_{k-1}\| < \|x_k - x_0\| + \frac{\rho - \|x_k - x_0\|}{\|x_k - x_{k-1}\|} \|x_k - x_{k-1}\| = \rho$$

(if $x_{k-1} = x_k$ holds, then (5a) implies $w_k = x_k \in B_\rho(x_0)$).

b) $\sum_{k\geq 0} \alpha_k \|x_k - x_{k-1}\|^2$ is summable since by $\alpha_k \leq \theta_k / \|x_k - x_{k-1}\|^2$ it holds that $\alpha_k \|x_k - x_{k-1}\|^2 \leq \theta_k$ and $\sum_{k\geq 0} \theta_k$ is summable by assumption.

In the next proposition we compare the squared distances $\|w_k - x^*\|^2$ and $\|x_{k+1} - x^*\|^2$, where x^* is any solution of $F(x) = y$ inside the ball $B_\rho(x_0)$.

prop:gain **Proposition 2.6.** *Let (A1) – (A4) hold and (x_k), (w_k) be sequences generated by Algorithm 1 (with (λ_k) and (α_k) chosen as in Steps [1] and [2.3] respectively). If $w_k \in B_\rho(x_0)$ then*

$$\|w_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \ \geq\ \|w_k - x_{k+1}\|^2 + 2(q-\eta)\lambda_k^{-1} \|D_k\| \|F(w_k) - y\|\,,\ k \geq 0$$

for any $x^* \in B_\rho(x_0)$ solution of $F(x) = y$,

Proof. From Lemma 2.3 (b) follows

$$\begin{aligned} \|w_k - x\|^2 - \|x_{k+1} - x\|^2 &= \|w_k - x_{k+1}\|^2 + 2\langle w_k - x_{k+1}, x_{k+1} - x\rangle \\ &= \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \langle A_k^* [F(w_k) + A_k(x_{k+1} - w_k) - y], x_{k+1} - x\rangle \\ &= \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \langle D_k, A_k(x_{k+1} - x)\rangle \\ &= \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \langle D_k, A_k\,(x_{k+1} - x \pm w_k) \pm F(w_k) \pm y\rangle \\ &= \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \left[\|D_k\|^2 + \langle D_k, A_k\,(w_k - x) - F(w_k) + y)\rangle \right], \end{aligned}$$

for $x \in X$ and $k \geq 0$. From this equation with $x = x^*$, (A2) and Lemma 2.3 (c) follows

$$\begin{aligned} \|w_k - x\|^2 - \|x_{k+1} - x\|^2 &= \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \left[\|D_k\|^2 + \langle D_k, F(x^*) - F(w_k) - A_k\,(x^* - w_k)\rangle \right] \\ &\geq \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \|D_k\| \Big[\|D_k\| - \eta \|y - F(w_k)\| \Big] \\ &\geq \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1} \|D_k\|\,(q-\eta)\,\|y - F(w_k)\|\,, \end{aligned}$$

completing the proof. □

In the following proposition, we examine the boundedness of the sequences (x_k) and (w_k) generated by Algorithm 1.

prop:bound **Proposition 2.7.** *Let (A1) – (A4) hold and (x_k), (w_k) be sequences generated by Algorithm 1 (with (λ_k) and (α_k) chosen as in Steps [1] and [2.3] respectively). If (α_k) satisfies Assumption 2.4 then (x_k) and (w_k) are contained in $B_\rho(x_0)$.*

Proof. We present here a proof by induction, with inductive step stated as follows:

Assume that $(w_k)_{k=0}^{l-1}$, $(x_k)_{k=0}^{l} \in B_\rho(x_0)$, and conclude that w_l, $x_{l+1} \in B_\rho(x_0)$.

For $l = 1$, it holds $w_0 = x_0 \in B_\rho(x_0)$ and $x_1 := w_0 + s_0$, where $\big(A_0^* A_0 + \lambda_0 I\big) s_0 = A_0^*[y - F(w_0)]$. Thus, from Proposition 2.6, (A3), (A4) follows $\|w_0 - x^\star\| \geq \|x_1 - x^\star\|$, hence $x_1 \in B_\rho(x_0)$. Now, the fact $x_0, x_1 \in B_\rho(x_0)$ together with Assumption 2.4 imply $w_1 \in B_\rho(x_0)$ (see Remark 2.5)).

For $l > 1$, the inductive assumption ensures x_{l-1}, $x_l \in B_\rho(x_0)$. Consequently, Assumption 2.4 implies $w_l \in B_\rho(x_0)$. Thus, arguing with Proposition 2.6, (A3), (A4) and the inductive assumption we obtain $\|w_l - x^\star\| \geq \|x_{l+1} - x^\star\|$. Therefore, $x_{l+1} \in B_\rho(x_0)$, concluding the proof. □

In the upcoming proposition we discuss the summability of three series related to inLM, a crucial element for proving a convergence theorem (see Theorem 2.9).

prop:series **Proposition 2.8.** *Assume that α_k fulfills Assumption 2.4, that (A1)-(A4) are fulfilled and that (x_k) and (w_k) are generated by Algorithm 1. Then it holds that the limit $\|x_k - x^*\|$ exist for all solutions x^* and that*

$$\sum_{k=0}^{\infty} \|x_{k+1} - w_k\|^2 < \infty\,, \quad \sum_{k=0}^{\infty} \lambda_k^{-1} \|F(w_k) - y\|^2 < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \lambda_k^{-1} \|F(x_k) - y\|^2 < \infty.$$

Proof. From (7) with $x = x^\star$ and Proposition 2.6 we conclude that[1]

$$\begin{aligned}(1+\alpha_k)\|x_k - x^*\|^2 - \alpha_k \|x_{k-1} - x^*\|^2 + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 - \|x_{k+1} - x^*\|^2 = \\ = \|w_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \ \geq \ \|w_k - x_{k+1}\|^2 + 2\lambda_k^{-1}(q-\eta)\,\|D_k\|\,\|F(w_k) - y\|\,,\ k \geq 0\,.\end{aligned}$$

Thus, defining $\varphi_k := \|x_k - x^*\|^2$ and $\eta_k := \alpha_k \|x_k - x_{k-1}\|^2$, we obtain

$$\begin{aligned}\alpha_k(\varphi_k - \varphi_{k-1}) - (\varphi_{k+1} - \varphi_k) + (1+\alpha_k)\eta_k \ \geq \\ \geq \ \|x_{k+1} - w_k\|^2 + 2\lambda_k^{-1}(q-\eta)\,\|D_k\|\,\|F(w_k) - y\|\,,\ k \geq 0\,. \end{aligned} \tag{9}$$ eq:gamma

Since $\alpha_k < 1$ we get from there that

$$\begin{aligned}\varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) &\leq (1+\alpha_k)\eta_k - \|w_k - x_{k+1}\|^2 - \tfrac{2(q-\eta)}{\lambda_k}\|D_k\|\|F(w_k) - y\| \\ &\leq 2\eta_k.\end{aligned}$$

We abbreviate $\zeta_k := \varphi_k - \varphi_{k-1}$ and write $[\zeta_k]_+$ for the positive part and get

$$\zeta_{k+1} \leq \alpha_k[\zeta_k]_+ + 2\eta_k$$

and hence with $\alpha_k \leq \alpha < 1$

$$[\zeta_{k+1}]_+ \leq \alpha_k[\zeta_k]_+ + 2\eta_k \leq \cdots \leq \alpha^k[\zeta_1]_+ + 2\sum_{j=0}^{k-1} \alpha^j \eta_{k-j}.$$

We sum this inequality from $k = 0, \ldots, \infty$ and get

$$\sum_{k=0}^{\infty} [\zeta_{k+1}]_+ \leq \tfrac{1}{1-\alpha}[\zeta_1]_+ + 2\sum_{k=0}^{\infty}\sum_{j=0}^{k-1} \alpha^j \eta_{k-j}\,. \tag{10}$$ eq:sum-th

[1]Notice that in Algorithm 1 we define x_k for $k \geq 0$ and α_k for $k \geq 1$. For this proof we additionally define $x_{-1} := x_0$ and $\alpha_0 := 0$; thus, (7) holds trivially for $k = 0$.

The latter sum can be calculated by swapping the order and substitution:

$$\sum_{k=0}^{\infty}\sum_{j=0}^{k-1}\alpha^j\eta_{k-j} = \sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\alpha^j\eta_{k-j}$$
$$\sum_{j=0}^{\infty}\alpha^j\sum_{l=1}^{\infty}\eta_l = \tfrac{1}{1-\alpha}\sum_{l=1}^{\infty}\eta_l\,.$$

Thus (10) turns into

$$\sum_{k=0}^{\infty}[\zeta_{k+1}]_+ \leq \tfrac{1}{1-\alpha}\left([\zeta_1]_+ + 2\sum_{l=1}^{\infty}\eta_l\right)\,.$$

The series on the right hand side is convergent by assumption and hence $\sum_{k=0}^{\infty}[\zeta_k]_+ < \infty$.

Now we define $\gamma_k = \varphi_k - \sum_{j=1}^{k}[\zeta_j]_+$ which is bounded from below since $\varphi_k \geq 0$ and the series is convergent. Moreover we have (recalling the definition of ζ_k)

$$\begin{aligned}\gamma_{k+1} &= \varphi_{k+1} - [\zeta_{k+1}]_+ - \sum_{j=1}^{k}[\zeta_j]_+ \\ &\leq \varphi_{k+1} - \varphi_{k+1} + \varphi_k - \sum_{j=1}^{k}[\zeta_j]_+ = \gamma_k\,.\end{aligned}$$

We see that γ_k is non-increasing and bounded from below, hence convergent. This implies that the limits

$$\lim_{k\to\infty}\varphi_k = \lim_{k\to\infty}\gamma_k + \sum_{j=1}^{\infty}[\zeta_j]_+$$

all exist and since $\varphi_k = \|x_k - x^*\|^2$ we get that $\lim_{k\to\infty}\|x_k - x^*\|$ exists.

Now we start at (9) again and write it as

$$2\lambda_k^{-1}(q-\eta)\|D_k\|\,\|F(w_k)-y\| + \|x_{k+1}-w_k\|^2 \leq 2\eta_k + \varphi_k - \varphi_{k+1} + \alpha_k[\zeta_k]_+\,.$$

Using Lemma 2.3 c) we get

$$2\lambda_k^{-1}q(q-\eta)\|F(w_k)-y\|^2 + \|x_{k+1}-w_k\|^2 \leq 2\eta_k + \varphi_k - \varphi_{k+1} + \alpha_k[\zeta_k]_+\,.$$

Summing from $k = 0, \ldots, \infty$ gives

$$\sum_{k=0}^{\infty}\left(2\lambda_k^{-1}q(q-\eta)\|F(w_k)-y\|^2 + \|x_{k+1}-w_k\|^2\right) \leq \varphi_1 + \sum_{k=0}^{\infty}(\alpha_k[\zeta_k]_+ + 2\eta_k)$$

and since the right hand side of this inequality is bounded, we get that $\sum_k\|x_{k+1}-w_k\|^2 < \infty$ and $\sum_k\lambda_k^{-1}\|F(w_k)-y\|^2 < \infty$. Lemma 2.3 d) implies that the last series is convergent as well. □

ssec:2.2

2.2 A strong convergence result

In what follows we prove a (strong) convergence result for the inLM method (Algorithm 1) in the exact data case. To prove this result we use two additional assumptions:

(A5) There exists $\lambda_{max} > 0$ s.t. $\lambda_k \leq \lambda_{max}$, for $k \geq 0$;

(A6) (α_k) is monotone non-increasing (see Step [2.3] of Algorithm 1).

th:conv **Theorem 2.9** (Convergence). *Let (A1) – (A6) hold and (x_k), (w_k) be sequences generated by Algorithm 1 (with (λ_k) and (α_k) chosen as in Steps [1] and [2.3] respectively). Additionally, assume that (α_k) complies with Assumption 2.4. Then, either the sequences (x_k), (w_k) stop after a finite number $k_0 \in \mathbb{N}$ of steps (in this case it holds $x_{k_0+1} = w_{k_0}$ and $F(w_{k_0}) = y$), or there exists $\bar{x} \in B_\rho(x_0)$, solution of $F(x) = y$, s.t. $\lim_k x_k = \lim_k w_k = \bar{x}$.*

Proof. We consider two cases.

Case I: $F(w_{k_0}) = y$ for some $k_0 \in \mathbb{N}$.
In this case, the sequences (x_k), (w_k) read $x_0, \dots, x_{k_0+1}$ and $w_0, \dots, w_{k_0}$. Moreover, it holds $x_{k_0+1} = w_{k_0}$ and $F(w_{k_0}) = y$ (see Remark 2.1).

Case II: $F(w_k) \neq y$, for every $k \geq 0$.
Notice that, in this case, the real sequence $\big(\|F(x_k) - y\|\big)$ is strictly positive. Moreover, it follows from (A5) and Proposition 2.8 (see second series) that $\lim_k \|F(w_k) - y\| = 0$. Therefore, there exists a strictly monotone increasing sequence $(l_j) \in \mathbb{N}$ satisfying

$$\|F(w_{l_j}) - y\| \;\leq\; \|F(w_k) - y\| \,, \quad \text{for} \quad k = 0, \dots, l_j \,. \tag{11}$$ eq:min

Notice that, given $k > 0$ and $z \in B_\rho(x_0)$, it holds

$$\begin{aligned}
\|w_k - z\|^2 - \|x_{k+1} - z\|^2 &= -\|x_{k+1} - w_k\|^2 - 2\langle x_{k+1} - w_k, w_k - z\rangle \\
&\leq 2\langle w_k - x_{k+1}, w_k - z\rangle \\
&= 2\langle \lambda_k^{-1} A_k^*(F(w_k) + A_k(x_{k+1} - w_k) - y), w_k - z\rangle \\
&= 2\lambda_k^{-1}\langle -F(x_{k+1}) + F(w_k) + A_k(x_{k+1} - w_k) + F(x_{k+1}) - y, A_k(w_k - z)\rangle \\
&\leq 2\lambda_k^{-1}\Big[\eta\|F(x_{k+1}) - F(w_k)\| + \|F(x_{k+1}) - y\|\Big]\, \|A_k(w_k - z)\| \\
&\leq 2\lambda_k^{-1}(1+\eta)\Big[\eta\|F(x_{k+1}) - F(w_k) \pm y\| + \|F(x_{k+1}) - y\|\Big]\, \|F(w_k) - F(z) \pm y\| \\
&\leq 2\lambda_k^{-1}(1+\eta)^2\Big[\|F(x_{k+1}) - y\|\|F(w_k) - y\| + \|F(x_{k+1}) - y\|\|F(z) - y\|\Big] \\
&\quad +2\lambda_k^{-1}\eta(1+\eta)\Big[\|F(w_k) - y\|^2 + \|F(w_k) - y\|\|F(z) - y\|\Big]
\end{aligned}$$

(in the second inequality we used (A2); in the third inequality we used [10, Eq.(11.7)]). Taking $z = w_{l_j}$ in the last inequality and arguing with Lemma 2.3 (d) and (11) it follows

$$\|w_k - w_{l_j}\|^2 - \|x_{k+1} - w_{l_j}\|^2 \;\leq\; 2(1+\eta)\Big[2\frac{(1+\eta)^2}{1-\eta} + 2\eta\Big]\, \lambda_k^{-1}\|F(w_k) - y\|^2 \;=:\; \mu_k \,, \tag{12}$$ eq:muk

for $k = 0, \dots, l_j$. Next we estimate the second term on the left-hand-side of (12). Lemma 2.2 (with $x = w_{l_j}$) implies

$$\|w_{k+1} - w_{l_j}\|^2 = (1 + \alpha_{k+1})\|x_{k+1} - w_{l_j}\|^2 - \alpha_{k+1}\|x_k - w_{l_j}\|^2 + \alpha_{k+1}(1 + \alpha_{k+1})\|x_{k+1} - x_k\|^2,$$

for $k = 0, 1, \dots$; from where we conclude that

$$0 \leq \|x_{k+1} - w_{l_j}\|^2 \;\leq\; \|w_{k+1} - w_{l_j}\|^2 + \alpha_{k+1}\big(\|x_k - w_{l_j}\|^2 - \|x_{k+1} - w_{l_j}\|^2\big) \,,\ k \geq 0 \,. \tag{13}$$ eq:xlj

Now, combining (12) with (13), and arguing with (A6), we obtain

$$\begin{aligned}
\|w_k - w_{l_j}\|^2 - \|w_{k+1} - w_{l_j}\|^2 &\leq \alpha_{k+1}\big(\|x_k - w_{l_j}\|^2 - \|x_{k+1} - w_{l_j}\|^2\big) + \mu_k \\
&\leq \alpha_k\|x_k - w_{l_j}\|^2 - \alpha_{k+1}\|x_{k+1} - w_{l_j}\|^2 + \mu_k \,, \qquad (14)
\end{aligned}$$

for $k = 0, \ldots, l_j$. Let $0 \le m \le l_j$. Adding up (14) for $k = m, \ldots, l_j - 1$ follows

$$\|w_m - w_{l_j}\|^2 - \|w_{l_j} - w_{l_j}\|^2 \ \le \ \alpha_m \|x_m - w_{l_j}\|^2 - \alpha_{l_j}\|x_{l_j} - w_{l_j}\|^2 + \sum_{k=m}^{l_j-1} \mu_k \,,$$

from where we derive that, for any $\varepsilon > 0$ it holds

$$\begin{aligned} \|w_m - w_{l_j}\|^2 \ &\le \ \alpha_m \|x_m - w_{l_j} \pm w_m\|^2 + \sum_{k=m}^{l_j-1} \mu_k \\ &\le \ \alpha_m \big((1+\varepsilon)\|x_m - w_m\|^2 + (1+\tfrac{1}{\varepsilon})\|w_m - w_{l_j}\|^2 \big) + \sum_{k=m}^{l_j} \mu_k \\ &\le \ (1+\varepsilon) \sum_{k=m}^{l_j} \alpha_k \|x_k - w_k\|^2 + (1+\tfrac{1}{\varepsilon})\alpha_m \|w_m - w_{l_j}\|^2 + \sum_{k=m}^{l_j} \mu_k \,. \end{aligned}$$

Consequently, whenever $m \le l_j - 1$, it holds

$$(1 - \tfrac{\varepsilon+1}{\varepsilon}\alpha_m)\|w_m - w_{l_j}\|^2 \ \le \ (1+\varepsilon) \sum_{k=m}^{l_j} \alpha_k \|x_k - w_k\|^2 + \sum_{k=m}^{l_j} \mu_k \,.$$

Now we choose $\varepsilon > 0$ such that $\frac{\varepsilon+1}{\varepsilon}\alpha_0 < 1$, define $\beta := (1 - \frac{\varepsilon+1}{\varepsilon}\alpha_0)^{-1}$, it follows from (A6) and (5a)

$$\begin{aligned} \|w_m - w_{l_j}\|^2 \ &\le \ 2\beta \sum_{k=m}^{\infty} \alpha_k \|x_k - w_k\|^2 + \beta \sum_{k=m}^{\infty} \mu_k \\ &\le \ 2\beta\alpha^3 \sum_{k=m}^{\infty} \|x_k - x_{k-1}\|^2 + \beta \sum_{k=m}^{\infty} \mu_k \,, \ m < l_j \end{aligned} \tag{15}$$

(notice that $\beta > 0$ and $\alpha_k \le \alpha$ due to (A6)).

Notice that (A2) together with Proposition 2.8 guarantee the summability of both series $\sum_k \mu_k$ and $\sum_k \|x_k - x_{k-1}\|^2$. Thus, defining $s_m := 2\beta\alpha^3 \sum_{k\ge m} \|x_k - x_{k-1}\|^2 + \beta \sum_{k \ge m} \mu_k$, for $m \in \mathbb{N}$, follows $s_m \to 0$ as $m \to \infty$.

Let $n > m$ be given. Choosing $l_j > n$, it follows from (15)

$$\|w_n - w_m\| \ \le \ \|w_n - w_{l_j}\| + \|w_{l_j} - w_m\| \ \le \ \sqrt{s_n} + \sqrt{s_m} \ \le \ 2\sqrt{s_m} \,,$$

from where we conclude that (w_k) is a Cauchy sequence. Consequently, (w_k) converges to some $\bar{x} \in X$. From Proposition 2.8 (see first series) it follows $\lim_k x_k = \lim_k w_k = \bar{x}$.

It remais to prove that $\bar{x}$ is a solution of $F(x) = y$. It suffices to verify that $\|F(w_k) - y\| \to 0$ as $k \to \infty$. This fact, however, is a consequence of Proposition 2.8 (see second series) together with Assumption (A5). □

2.3 Regularization properties

ssec:2.3

In this section we address the noisy data case, i.e. $\delta > 0$, and investigate regularization properties of the inertial Levenberg-Marquardt method. For noisy data the inLM method reads is stated in Algorithm 2.

rem:alg2

Remark 2.10 (Comments regarding Algorithm 2).
The discrepancy principle *is used as stopping criterion in Algorithm 2. i.e. the loop in Step [2] terminates at step $k^* = k^*(\delta, y^\delta)$ s.t. $k^* := \min\{k \in \mathbb{N};\ \|F(w_k^\delta) - y^\delta\| \le \tau\delta\}$, where $\tau > 1$.*

Note that Algorithm 2 generate sequences $(x_k^\delta)_{k=0}^{k^*+1}$ and $(w_k^\delta)_{k=0}^{k^*}$. The finiteness of the stopping index k^* in Step [2.5] is addressed in Proposition 2.13.

[0] choose an initial guess $x_0 \in X$; set $w_0^\delta := x_0$; $k := 0$; flag := 'FALSE';

[1] choose $\tau > (\eta+1)(q-\eta)^{-1}$, $\alpha \in [0,1)$ and $(\lambda_k)_{k\geq 0} \in \mathbb{R}^+$;

[2] **repeat**

if $\big(\|F(w_k^\delta) - y^\delta\| > \tau\delta\big)$ **then**

[2.1] $A_k^\delta := F'(w_k^\delta)$; compute $s_k^\delta \in X$ as the solution of
$$\big((A_k^\delta)^* A_k^\delta + \lambda_k I\big)\, s_k^\delta \;=\; (A_k^\delta)^* \big(y^\delta - F(w_k^\delta)\big);$$

[2.2] $x_{k+1}^\delta := w_k^\delta + s_k^\delta$;

[2.3] $k := k+1$;

[2.4] choose $\alpha_k^\delta \in [0,\alpha]$; $w_k^\delta := x_k^\delta + \alpha_k^\delta(x_k^\delta - x_{k-1}^\delta)$;

else

[2.5] $s_k^\delta := 0$; $x_{k+1}^\delta := w_k^\delta$; $k^* := k$; flag := 'TRUE';

end if

until $\big($flag = 'TRUE'$\big)$

alg:init-noise

Algorithm 2: Inertial Levenberg-Marquardt method in the noisy data case.

For each $0 \leq k \leq k^*$, define $D_k^\delta := F(w_k^\delta) + A_k^\delta(x_{k+1}^\delta - w_k^\delta) - y^\delta$. It is straightforward to verify that the results stated in Lemma 2.2 and Lemma 2.3 remain valid in the noisy data case (the corresponding proofs are analogous and will be omitted).

Additionally, if the sequence of inertial paramenters (α_k^δ) in Algorithm 2 is chosen in accordance with

$$\alpha_k^\delta := \begin{cases} \min\left\{ \dfrac{\theta_k}{\|x_k^\delta - x_{k-1}^\delta\|^2}, \dfrac{\min\{\theta_k,\ \rho - \|x_k^\delta - x_0\|\}}{\|x_k^\delta - x_{k-1}^\delta\|}, \alpha \right\}, & \text{if } \|x_k^\delta - x_{k-1}^\delta\| > 0 \\ 0 & \text{, otherwise} \end{cases} \tag{16}$$

def:alpha

(where (θ_k) is chosen as in Assumption 2.4), then Remark 2.5 (a) holds true for $k = 1, \ldots, k^*$.

In the sequel we extend the "gain estimate" in Proposition 2.6 to the noisy data case.

prop:gainN

Proposition 2.11. *Let (A1) – (A4) hold and (x_k^δ), (w_k^δ) be sequences generated by Algorithm 2 (with (λ_k) and (α_k^δ) chosen as in Steps [1] and [2.4] respectively). If $w_k^\delta \in B_\rho(x_0)$ for some $0 \leq k \leq k^*$, then*

$$\|w_k^\delta - x^*\|^2 - \|x_{k+1}^\delta - x^*\|^2 \geq \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\|D_k^\delta\|\big[(q-\eta)\|y^\delta - F(w_k^\delta)\| - (\eta+1)\delta\big],$$

for any $x^* \in B_\rho(x_0)$ solution of $F(x) = y$.

Proof. Since Lemma 2.3 remains valid in the noise data case (see Remark 2.10), we make a similar argument as in the proof of Proposition 2.6 to establish that

$$
\begin{aligned}
\|w_k^\delta - x^*\|^2 - \|x_{k+1}^\delta - x^*\|^2 &= \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\langle w_k^\delta - x_{k+1}^\delta, x_{k+1}^\delta - x^*\rangle \\
&= \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\big\langle (A_k^\delta)^*\big[F(w_k^\delta) + A_k^\delta(x_{k+1}^\delta - w_k^\delta) - y^\delta\big], x_{k+1}^\delta - x^*\big\rangle \\
&= \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\big\langle D_k^\delta, A_k^\delta(x_{k+1}^\delta - x^*)\big\rangle \\
&= \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\big\langle D_k^\delta, A_k^\delta(x_{k+1}^\delta - x^*) \pm A_k^\delta w_k^\delta \pm F(w_k^\delta) \pm y^\delta\big\rangle \\
&= \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\big[\|D_k^\delta\|^2 + \big\langle D_k^\delta, A_k^\delta(w_k^\delta - x^*) - F(w_k^\delta) + y^\delta\big\rangle\big] \\
&= \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\big[\|D_k^\delta\|^2 + \big\langle D_k^\delta, F(x^*) - F(w_k^\delta) - A_k^\delta(x^* - w_k^\delta) + y^\delta - y\big\rangle\big] \\
&\geq \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\big[\|D_k^\delta\|^2 - \eta\|D_k\|\|y - F(w_k^\delta)\| - \|D_k^\delta\|\delta\big] \\
&\geq \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\|D_k^\delta\|\big[\|D_k^\delta\| - \eta\|y^\delta - F(w_k^\delta)\| - (\eta+1)\delta\big] \\
&\geq \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\|D_k^\delta\|\big[(q-\eta)\|y^\delta - F(w_k^\delta)\| - (\eta+1)\delta\big],
\end{aligned}
$$

completing the proof. □

cor:boundN **Corollary 2.12.** *Due to the choice $\tau > (\eta+1)(q-\eta)^{-1}$ in Step [1] of Algorithm 2, it follows from Proposition 2.11*

$$\|w_k^\delta - x^\star\|^2 - \|x_{k+1}^\delta - x^\star\|^2 \;\geq\; \|w_k^\delta - x_{k+1}^\delta\|^2 + 2\lambda_k^{-1}\|D_k^\delta\|\big[(q-\eta)\tau\delta - (\eta+1)\delta\big] \;\geq\; 0,$$

for $k = 0, \ldots k^*$. Consequently, under the assumptions of Proposition 2.11, if (α_k^δ) satisfies (16) then x_k^δ, $w_k^\delta \in B_\rho(x_0)$ for $k = 0, \ldots k^*$ (the proof of this assertion follows the lines of the proof of Proposition 2.7 and is omitted).

In the sequel we address the finiteness of the stopping index k^* as defined in Step [2.5] of Algorithm 2.

prop:kstar **Proposition 2.13.** *Let (A1) – (A4) hold and (x_k^δ), (w_k^δ) be sequences generated by Algorithm 2 (with (λ_k) and (α_k^δ) chosen as in Steps [1] and [2.4] respectively). Assume that (α_k^δ) satisfies (16). If $\sum_k \lambda_k^{-1} = \infty$ the stopping index k^* defined in Step [2.5] is finite. Additionaly, if $\lambda_k \leq \lambda_{\max}$ then*

$$k^* \;\leq\; \lambda_{\max}\Big(2q\tau\delta^2\big[(q-\eta)\tau - (\eta+1)\big]\Big)^{-1}\Big[\rho^2 + 2\textstyle\sum_k \theta_k\Big].$$

Proof. Recall that Lemma 2.2 and Lemma 2.3 remain valid in the noisy data case (see Remark 2.10). We claim that, if k^* is not finite the sequence of partial sums (σ_n) defined by $\sigma_n := \sum_{k=0}^n \alpha_k^\delta\big(\|x_{k-1}^\delta - x^\star\|^2 - \|x_k^\delta - x^\star\|^2\big)$ is bounded (here $x_{-1}^\delta := x_0^\delta$). Indeed, from Lemma 2.2 (with $x = x^\star$) and Proposition 2.11 follow

$$
\begin{aligned}
(1+\alpha_k^\delta)\|x_k^\delta - x^\star\|^2 - \alpha_k^\delta\|x_{k-1}^\delta - x^\star\|^2 + \alpha_k^\delta(1+\alpha_k^\delta)\|x_k^\delta - x_{k-1}^\delta\|^2 - \|x_{k+1} - x^\star\|^2 \;= \\
= \|w_k^\delta - x^\star\|^2 - \|x_{k+1}^\delta - x^\star\|^2 \;\geq\; 0\,,\;\; k \geq 0\,.
\end{aligned}
$$

Thus, $\alpha_k^\delta\big(\|x_k^\delta - x^\star\|^2 - \|x_{k-1}^\delta - x^\star\|^2\big) + \|x_k^\delta - x^\star\|^2 + 2\alpha_k^\delta\|x_k^\delta - x_{k-1}^\delta\|^2 - \|x_{k+1}^\delta - x^\star\|^2 \geq 0$. Consequently, $\alpha_k^\delta\big(\|x_{k-1}^\delta - x^\star\|^2 - \|x_k^\delta - x^\star\|^2\big) \leq \|x_k^\delta - x^\star\|^2 - \|x_{k+1}^\delta - x^\star\|^2 + 2\alpha_k^\delta\|x_k^\delta - x_{k-1}^\delta\|^2$, for $k \geq 0$. Adding the last inequality for $k = 0, \ldots n$, and using (16) we obtain

$$\sigma_n \;\leq\; \|x_0^\delta - x^\star\|^2 - \|x_{n+1}^\delta - x^\star\|^2 + 2\sum_{k=0}^n \alpha_k^\delta\|x_k^\delta - x_{k-1}^\delta\|^2 \;\leq\; \rho^2 + 2\sum_{k=0}^\infty \theta_k\,. \qquad (17)$$ eq:sigmaN

The boundedness of sequence (σ_n) follows from the summability of (θ_k), proving our claim.

For each $0 \leq k \leq k^*$ we derive from Proposition 2.11 and Lemma 2.3 (c)

$$
\begin{aligned}
2\lambda_k^{-1} q\tau\delta^2[(q-\eta)\tau - (\eta+1)] &\leq 2\lambda_k^{-1}\|D_k^\delta\|\big[(q-\eta)\|y^\delta - F(w_k^\delta)\| - (\eta+1)\delta\big] \\
&\leq \|w_k^\delta - x^\star\|^2 - \|x_{k+1}^\delta - x^\star\|^2.
\end{aligned}
$$

This inequality, Lemma 2.2 (with $x = x^\star$), (16) and Corollary 2.12 allow us to conclude that

$$\begin{aligned} 2\lambda_k^{-1} q\tau\delta^2[(q-\eta)\tau - (\eta+1)] &\le \|x_k^\delta - x^\star\|^2 - \|x_{k+1}^\delta - x^\star\|^2 + \alpha_k^\delta \|x_k^\delta - x^\star\|^2 \\ &\quad - \alpha_k^\delta \|x_{k-1}^\delta - x^\star\|^2 + \alpha_k^\delta(1+\alpha_k^\delta)\|x_k^\delta - x_{k-1}^\delta\|^2 \\ &\le \|x_k^\delta - x^\star\|^2 - \|x_{k+1}^\delta - x^\star\|^2 + \alpha_k^\delta\big(\|x_k^\delta - x^\star\|^2 - \|x_{k-1}^\delta - x^\star\|^2\big) + 2\theta_k \end{aligned}$$

for $0 \le k \le k^*$. Summing up the last inequality for $k = 0, \dots, n$ with $n \le k^*$ gives us

$$2q\tau\delta^2\big[(q-\eta)\tau - (\eta+1)\big] \sum_{k=0}^{n} \lambda_k^{-1} \ \le\ \|x_0 - x^\star\|^2 + \sigma_n + 2\sum_{k=0}^{n} \theta_k\,. \tag{18}$$ eq:finite

If k^* is not finite, it follows from the boundedness of (σ_n) and the summability of (θ_k) that the right hand side of (18) is bounded. However, this contradicts the assumption $\sum_k \lambda_k^{-1} = \infty$. Thus, k^* has to be finite. To prove last assertion, note that the additional assumption $\lambda_k \le \lambda_{\max}$ together with (18) and (17) imply $2q\tau\delta^2\big[(q-\eta)\tau-(\eta+1)\big]\lambda_{\max}^{-1} k^* \le 2\rho^2 + 4\sum_k \theta_k$, concluding the proof. □

In the sequel we present the main results of this section, namely a stability result (see Theorem 2.14) and a semi-convergence result (see Theorem 2.15).

th:stabil **Theorem 2.14** (Stability). *Let (A1) hold, (δ^j) be a sequence of positive numbers converging to zero and (y^{δ^j}) be a sequence of noisy data satisfying $\|y^{\delta^j} - y\| \le \delta^j$, where $y \in Rg(F)$. For each $j \in \mathbb{N}$, let $(x_l^{\delta^j})_{l=0}^{k_j^*+1}$ and $(w_l^{\delta^j})_{l=0}^{k_j^*}$ be the corresponding sequences generated by Algorithm 2, with (λ_k) and $(\alpha_k^{\delta^j})$ chosen as in Steps [1] and [2.4] respectively (here k_j^* represent the stopping indices defined in Step [2.5]). Additionaly, assume that $(\alpha_k^{\delta^j})$ complies with (16).*

Let (x_k) and (w_k) be the sequences generated by Algorithm 1 with (α_k) satisfying Assumption 2.4. For each $k \ge 0$ it holds

$$\lim_{j\to\infty} x_k^{\delta^j} = x_k \quad \text{and} \quad \lim_{j\to\infty} w_k^{\delta^j} = w_k \tag{19}$$ eq:stabil

(in view of Remark 2.1, if (x_k) and (w_k) are finite then the first limit in (19) holds for $k = 0, \dots, k_0+1$, while the second limit holds for $k = 0, \dots, k_0$).[2]

Proof. We give an inductive proof. In what follows we adopt the simplifying notation $A_k^j := F'(w_k^{\delta^j})$. Notice that $w_0^{\delta^j} = w_0 = x_0 = x_0^{\delta^j}$ for all $j \in \mathbb{N}$. Thus, (19) holds for $k = 0$. Next, assume the existence of $(x_l)_{l\le k}$ and $(w_l)_{l\le k}$ generated by Algorithm 1 (and corresponding $(\alpha_l)_{l\le k}$ satisfying Assumption 2.4) such that $\lim_j x_l^{\delta^j} = x_l$ and $\lim_j w_l^{\delta^j} = w_l$, for $l = 0, \dots, k$.

If $F(w_k) = y$ then $k_0 := k$, $x_{k_0+1} = w_{k_0}$ and (19) holds only for a finite number of indexes (i.e. (x_k) and (w_k) are finite). If $F(w_k) \ne y$, it follows from Algorithms 1 and 2 that

$$s_k = (A_k^* A_k + \lambda_k I)^{-1} A_k^*(y - F(w_k)) \ \text{ and } \ s_k^{\delta^j} = [(A_k^j)^* A_k^j + \lambda_k I]^{-1}(A_k^j)^*(y^{\delta^j} - F(w_k^{\delta^j})).$$

Thus (A1), the assumption $\lim_j y^{\delta^j} = y$, the inductive assumption $\lim_j w_k^{\delta^j} = w_k$, and the fact $\min\big\{\|A_k^* A_k + \lambda_k I\|,\, \min_j \|(A_k^j)^* A_k^j + \lambda_k I\|\big\} \ge \lambda_k > 0$ allow us to conclude that $\lim_j s_k^{\delta^j} = s_k$. Consequently,

$$\lim_{j\to\infty} x_{k+1}^{\delta^j} \ =\ \lim_{j\to\infty}(w_k^{\delta^j} + s_k^{\delta^j}) \ =\ w_k + s_k \ =\ x_{k+1}\,. \tag{20}$$ eq:Lxkm1

At this point, two distinct cases must be considered:

Case I: $x_{k+1} \ne x_k$. Choose α_{k+1} according to Assumption 2.4. From $\lim_j x_k^{\delta^j} = x_k$, (20)

[2] In this case $x_{k_0+1} = w_{k_0}$ and $F(w_{k_0}) = y$.

and (16) it follows that $\lim_j \alpha_{k+1}^{\delta^j} = \alpha_{k+1}$. Defining w_{k+1} as in Step [2.3] of Algorithm 1, we conclude that

$$\lim_{j\to\infty} w_{k+1}^{\delta^j} \;=\; \lim_{j\to\infty} \big(x_{k+1}^{\delta^j} + \alpha_{k+1}^{\delta^j}(x_{k+1}^{\delta^j} - x_k^{\delta^j})\big) \;=\; x_{k+1} + \alpha_{k+1}(x_{k+1} - x_k) \;=\; w_{k+1} \,.$$

Case II: $x_{k+1} = x_k$. Assumption 2.4 implies $\alpha_{k+1} = 0$. Define w_{k+1} as in Step [2.1] of Algorithm 1 (in this case $w_{k+1} = x_{k+1}$). From $\lim_j x_k^{\delta^j} = x_k$, $\sup_j \alpha_{k+1}^{\delta^j} \le \alpha$, (20), and Step [2.4] of Algorithm 2, it follows that

$$\lim_{j\to\infty} w_{k+1}^{\delta^j} \;=\; \lim_{j\to\infty} \big[x_{k+1}^{\delta^j} + \alpha_{k+1}^{\delta^j}(x_{k+1}^{\delta^j} - x_k^{\delta^j})\big] \;=\; x_{k+1} \;=\; w_{k+1} \,.$$

Thus, in either case it holds $\lim_j w_{k+1}^{\delta^j} = w_{k+1}$, concluding the inductive proof. □

th:semi-conv **Theorem 2.15** (Semi-convergence). *Let (A1) - (A6) hold, (δ^j) be a sequence of positive numbers converging to zero, and (y^{δ^j}) be a sequence of noisy data satisfying $\|y^{\delta^j} - y\| \le \delta^j$, where $y \in Rg(F)$. For each $j \in \mathbb{N}$, let $(x_l^{\delta^j})_{l=0}^{k_j^*+1}$ and $(w_l^{\delta^j})_{l=0}^{k_j^*}$ be sequences generated by Algorithm 2, with (λ_l) and $(\alpha_l^{\delta^j})$ chosen as in Steps [1] and [2.4] respectively, and $(\alpha_l^{\delta^j})_{l=0}^{k_j^*}$ complying with (16) (here, $k_j^* = k^*(\delta^j, y^j)$ is the stopping index defined in Step [2.5]).*

The sequence $(x_{k_j^*}^{\delta^j})_j$ converges strongly to some $\bar{x} \in B_\rho(x_0)$, such that $F(\bar{x}) = y$.

Proof. Let (x_k), (w_k) be sequences generated by Algorithm 1 with exact data y and (α_k) satisfying Assumption 2.4. Since (A1) - (A6) hold, it follows from Theorem 2.9 the existence of $\bar{x} \in B_\rho(x_0)$, solution of $F(x) = y$, s.t. $\lim_k x_k = \lim_k w_k = \bar{x}$. We aim to prove that $\lim_j x_{k_j^*}^{\delta^j} = \bar{x}$. It suffices to prove that every subsequence of $(x_{k_j^*}^{\delta^j})_j$ has itself a subsequence converging strongly to $\bar{x}$.

Denote an arbitrary subsequence of $(x_{k_j^*}^{\delta^j})_j$ again by $(x_{k_j^*}^{\delta^j})_j$, and represent by $(k_j^*)_j \in \mathbb{N}$ the corresponding subsequence of indices. Two cases are considered:

Case 1. $(k_j^*)_j$ has a finite accumulation point.
In this case, we can extract a subsequence $(k_{j_m}^*)$ of (k_j^*) such that $k_{j_m}^* = n$, for some $n \in \mathbb{N}$ and all indices j_m. Applying Theorem 2.14 to (δ^{j_m}) and $(y^{\delta^{j_m}})$, we conclude that $w_{k_{j_m}^*}^{\delta^{j_m}} = w_n^{\delta^{j_m}} \to w_n$ and $x_{k_{j_m}^*+1}^{\delta^{j_m}} = x_{n+1}^{\delta^{j_m}} \to x_{n+1}$, as $j_m \to \infty$. We claim that $F(w_n) = y$. Indeed, $\|F(w_n) - y\| = \lim_{j_m} \|F(w_n^{\delta^{j_m}}) - y\| \le \lim_{j_m} \big(\|F(w_n^{\delta^{j_m}}) - y^{\delta^{j_m}}\| + \|y^{\delta^{j_m}} - y\|\big) \le \lim_{j_m} (\tau + 1)\, \delta^{j_m} = 0$. I.e. in this case, the second assertion of Theorem 2.9 holds with $k_0 = n$. Thus, $x_{n+1} = w_n = \bar{x}$.

Case 2. $(k_j^*)_j$ has no finite accumulation point.
In this case we can extract a monotone strictly increasing subsequence, again denoted by $(k_j^*)_j$. Take $\varepsilon > 0$. From Theorem 2.9 follows the existence of $K_1 = K_1(\varepsilon) \in \mathbb{N}$ such that

$$\|x_k - \bar{x}\| \;<\; \tfrac{1}{3}\varepsilon \,, \;\; k \ge K_1 \,. \tag{21}$$ eq:eps3-1

Since $\sum_k \theta_k$ is finite (see Assumption 2.4), there exists $K_2 = K_2(\varepsilon) \in \mathbb{N}$ such that

$$\sum_{k \ge K_2} \theta_k \;\le\; \tfrac{1}{3}\varepsilon \,. \tag{22}$$ eq:eps3-2

Define $K = K(\varepsilon) := \max\{K_1, K_2\}$. Due to the monotonicity of $(k_j^*)_j$, there exists $J_1 \in \mathbb{N}$ such that $k_j^* > K$ for $j \ge J_1$.

Theorem 2.14 applied to the subsequences $(\delta^j)_j$ and $(y^{\delta^j})_j$, corresponding to $(k_j^*)_j$, implies the existence of $J_2 \in \mathbb{N}$ s.t.

$$\|x_K^{\delta^j} - x_K\| \;\le\; \tfrac{1}{3}\varepsilon \,, \;\; j > J_2 \,. \tag{23}$$ eq:eps3-3

Set $J := \max\{J_1, J_2\}$. From Proposition 2.11 (with $x^* = \bar{x}$) and Step [2.4] of Algorithm 2 follow $\|x_{k+1}^{\delta j} - \bar{x}\| \le \|w_k^{\delta j} - \bar{x}\| \le \|x_k^{\delta j} - \bar{x}\| + \alpha_k^{\delta j} \|x_k^{\delta j} - x_{k-1}^{\delta j}\|$, for $j \ge J$ and $k = 0, \dots, k_j^* - 1$. Consequently,

$$\|x_{k+1}^{\delta j} - \bar{x}\| - \|x_k^{\delta j} - \bar{x}\| \ \le \ \alpha_k^{\delta j} \|x_k^{\delta j} - x_{k-1}^{\delta j}\| \,, \tag{24}$$

eq:telesc

for $j \ge J$ and $k = 0, \dots, k_j^* - 1$. Adding (24) for $k = K, \dots, k_j^* - 1$ we obtain

$$\|x_{k_j^*}^{\delta j} - \bar{x}\| \ \le \ \|x_K^{\delta j} - \bar{x}\| \ + \ \sum_{k=K}^{k_j^*-1} \alpha_k^{\delta j} \|x_k^{\delta j} - x_{k-1}^{\delta j}\| \,, \ j \ge J.$$

Thus, arguing with (16), together with (21), (22) and (23), we obtain

$$\|x_{k_j^*}^{\delta j} - \bar{x}\| \ \le \ \|x_K^{\delta j} - \bar{x}\| + \sum_{k=K}^{k_j^*-1} \theta_k \ \le \ \|x_K^{\delta j} - x_K\| + \|x_K - \bar{x}\| + \sum_{k \ge K} \theta_k \ \le \ \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon,$$

for $j \ge J$. Repeating the above argument for $\varepsilon = 1, \frac{1}{2}, \frac{1}{3}, \dots$ we generate a sequence of indices $j_1 < j_2 < j_3 < \dots$ such that $\|x_{k_{j_m}^*}^{\delta j_m} - \bar{x}\| \le \frac{1}{m}$, for $m \in \mathbb{N}$. This concludes Case 2, and completes the proof of the theorem. □

3 Numerical experiments

sec:numerics

In this section two distinct ill-posed problems are used to investigate the numerical efficiency of the inLM method.

3.1 Parameter identification in an elliptic PDE

We aim to identify the coefficient $c \ge 0$ in the elliptic PDE on the unit square $\Omega = (0,1)^2$ with Dirichlet boundary condition

$$-\Delta u + cu = g, \ \text{ in } \Omega \qquad u = \bar{u} \ \text{ on } \partial\Omega \tag{25}$$

eqn:PDE_r

from the knowledge of u on the full domain $S \subset \Omega$. Here the right-hand side $g \in L^2(\Omega)$ and boundary conditions $\bar{u} \in H^{3/2}(\Omega)$ are known. This is a typical benchmark inverse problem, see [14, Example 4.2]. If u has no zeroes in Ω, then c can be recovered explicitly by

$$c = (g + \Delta u)/u. \tag{26}$$

eqn:c_fro

However, for u given with noise, this operation is expected to be unstable as the application of the Laplacian is ill-conditioned. We rearrange and discretize (25) with a uniform grid of size $n \times n$ and use the standard five-point stencil Δ_n with fineness $1/n$ in both dimensions as a discretization of the Laplacian Δ. We associate functions on Ω with column vectors by assembling its values on the grid and traversing row-wise. We state the discretized inverse problem as asking for a reconstruction of $c^\delta \in \mathbb{R}^{n^2}$ from a given noisy solution $u^\delta \in \mathbb{R}^{n^2}$ and a vector $z \in \mathbb{R}^{n^2}$, which contains the discretized right-hand side g as well as boundary conditions $\bar{u}$ from (25), such that

$$F(c^\delta) := \big(- \Delta_n + \operatorname{diag}(c^\delta) \big)^{-1}(z) = u^\delta, \tag{27}$$

eqn:PDE_r

where $\operatorname{diag}(c^\delta)$ denotes the diagonal matrix with entries from c^δ. The mapping F from (27) is known to fulfill assumption (A2) locally. For g and $\bar{u}$ we set

$$g(x,y) = 200 \cdot e^{-10\left(x - \frac{1}{2}\right)^2 - 10\left(y - \frac{1}{2}\right)^2},$$
$$\bar{u}(x,y) = 0, \qquad (x,y) \in \partial\Omega$$

and $c^\dagger = c_0^\dagger * \varphi$ with

$$c_0^\dagger(x,y) = \begin{cases} 10, & \text{if } \min\left(\sqrt{(x-0.25)^2+(y-0.5)^2},\ \sqrt{(x-0.75)^2+(y-0.5)^2} \right) < \frac{1}{10}, \\ 0, & \text{otherwise}, \end{cases}$$

$$\varphi(x,y) = \frac{1}{20\sqrt{\pi}} \cdot e^{-\frac{x^2+y^2}{200}}, \qquad x, y \in (0,1)$$

and compute u^δ by a forward evaluation of the operator F from (27). We choose $n = 100$, which means that all vectors have size 10^4. We approximate the linear solve in step [2.1] of Algorithm 2 by two steps of the conjugate gradient method with initial value zero for the s-variable.

We first examine the noiseless case. We depict our chosen $c^\dagger$ and the corresponding solution $u^\dagger$ of the forward problem in Figure 1. The vector $u^\dagger$ is everywhere nonzero and hence we could recover $c^\dagger$ exactly by division as in (26). The rightmost subfigure in Figure 1 shows that this is perfectly possible in our case. Nevertheless, we test how well the iterates w^k of Algorithm 1 are able to approximate the coefficient $c^\dagger$. We choose $w_0 = 0$ and consider $\alpha = k/10$ for $k = 0, 2, ..., 10$, where $\alpha = 0$ corresponds to the non-accelerated Levenberg-Marquardt method and $\alpha = 1$ is not covered by our theory. To investigate convergence, we keep track of the residuals $\|F(w_k) - u^\dagger\|$ and the distances $\|c_k - c^\dagger\|$. The corresponding results can be seen in Figure 2. We observe that all methods converge, where convergence is faster for larger acceleration parameters α except for $\alpha = 1$. In Figure 3 we see that after 10 iterations, larger values of α proceed much faster in reconstruction and $\alpha = 1$ gives the best guess. Figure 4 shows that after 500 iterations the reconstructions look decent for $\alpha < 1$, but the peak shape is not fit for $\alpha = 1$.

Next, we add 1% of relative noise to the forward solution $u^\dagger$, which yields a noisy vector u^δ with no visible difference from $u^\dagger$ (left subfigures in Figure 1 and Figure 5). Here, the naive calculation of c^δ by (26) fails drastically, as one sees from the right subfigure in Figure 5. We compute reconstructions using Algorithm 2, where we again initialize by $w_0 = 0$ and choose $\alpha = k/10$ for $k = 0, 2, ..., 10$. Figure 6 shows the typical semi-convergence phenomenon. As one can see, the closest distance to the true coefficient is achieved earlier for larger values of α, which even includes $\alpha = 1$. We illustrate stopping by Morozov's discrepancy principle with the horizontal line in the right subfigure of Figure 6 and with bullet points on the graphs, where we set $\tau = 1$. From both Figure 6 (left subfigure) and from Figure 7 one can see that stopping happens too early even though we set $\tau = 1$. Indeed, we see that for $\alpha < 1$ the residuals decay even slightly below the absolute noise level, where the approached residual value does not depend of the concrete value of $\alpha < 1$. After 100 iterations, the reconstruction looks decent for $\alpha \le 0.8$, but breaks down for $\alpha = 1$, see Figure 8. In Figure 9 we see that the reconstruction at the respective iterations where w_k is closest to $c^\dagger$ in Euclidean norm do not look different for varying α (cf. Figure 6).

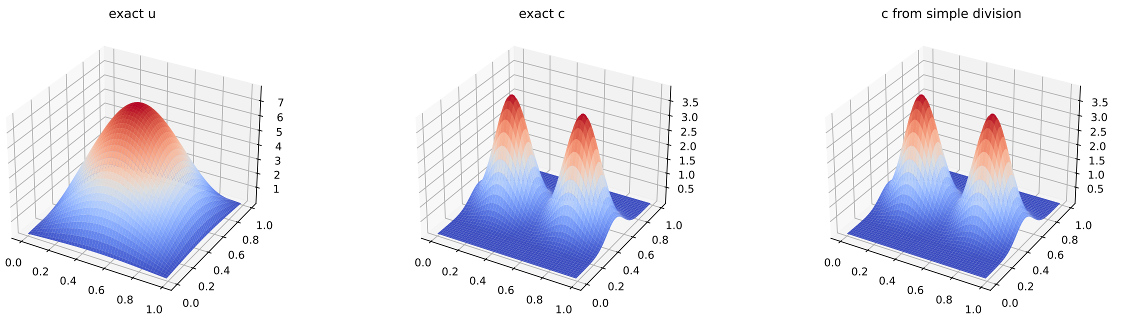

Figure 1: Left to right: exact PDE solution $u^\dagger$ (exact solution to the forward problem), $c^\dagger$ (exact solution to the inverse problem), reconstruction of $c^\dagger$ by (26) fig:noise

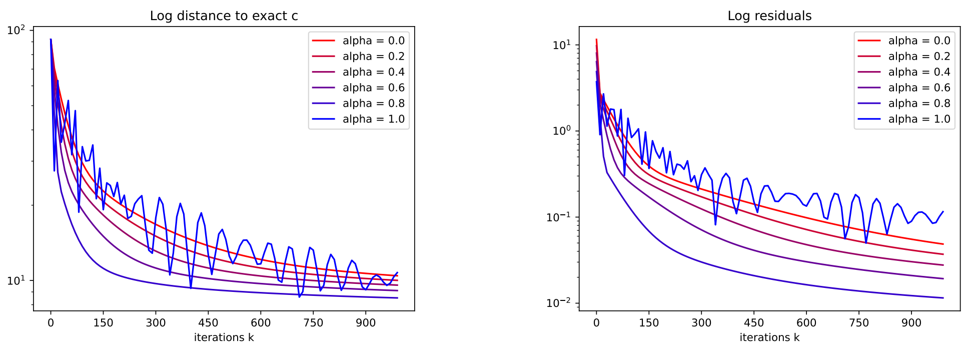

Figure 2: Results without noise: errors for $\alpha_k \equiv \alpha \in \{0, 0.2, ..., 1\}$ from red to blue color. Left: distances $\|c_k - c^\dagger\|_2$, right: residuals $\|F(c_k) - u^\dagger\|_2$ fig:noise

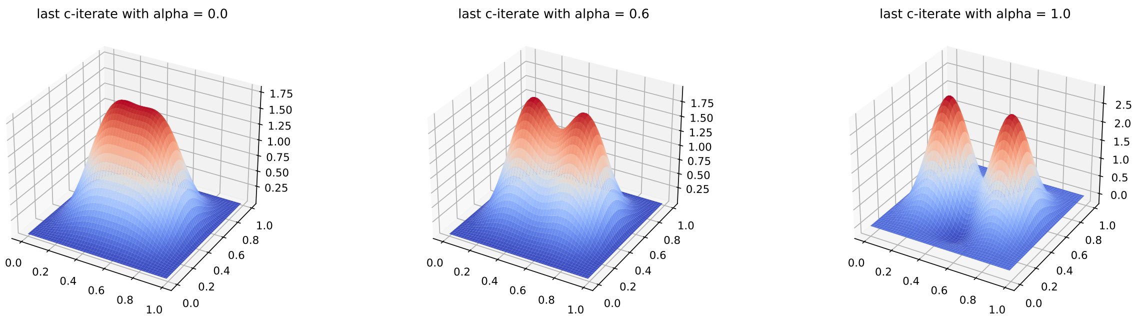

Figure 3: Results without noise, left to right: iterates w_{10} for $\alpha = 0$ (non-accelerated Levenberg-Marquardt method), $\alpha = 0.6$ and $\alpha = 1$ fig:noise

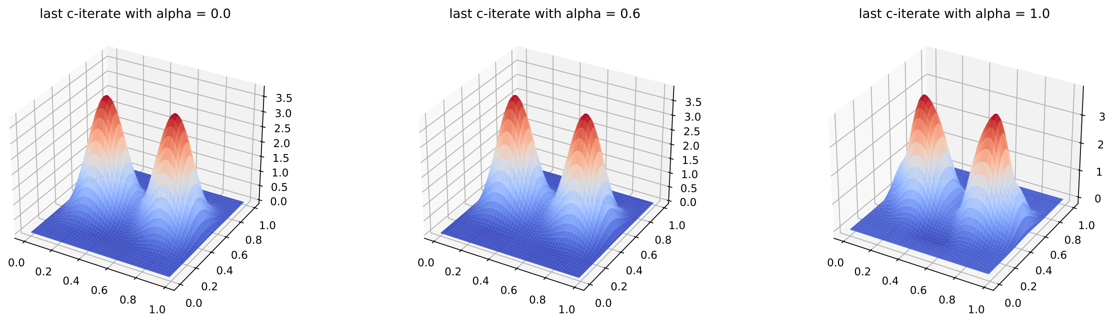

Figure 4: Results without noise, left to right: iterates w_{500} for $\alpha = 0$ (non-accelerated Levenberg-Marquardt method), $\alpha = 0.6$ and $\alpha = 1$ fig:noise

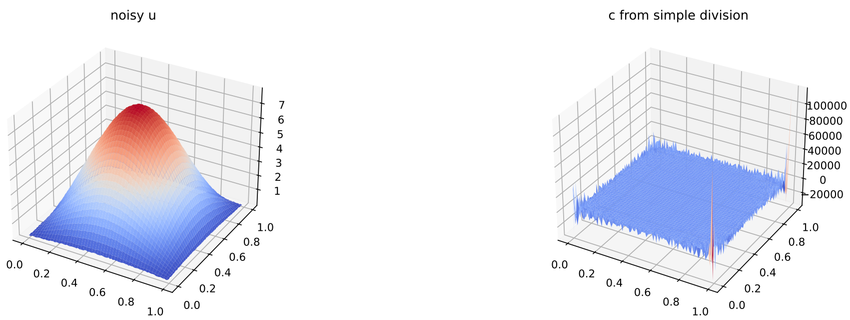

Figure 5: Left: noisy PDE solution u^δ, right: naive approximation of $c^\dagger$ by (26) fig:noisy

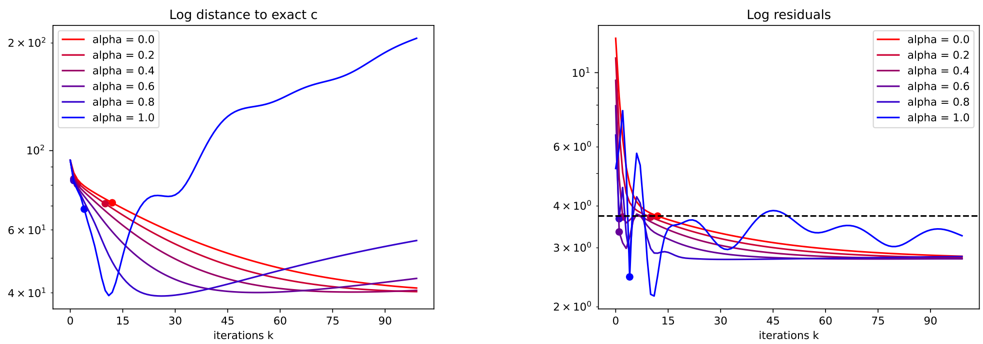

Figure 6: Results with 1% of relative noise, errors for $\alpha_k \equiv \alpha \in \{0, 0.2, ..., 1\}$ from red to blue color. Left: distances $\|c_k - c^\dagger\|_2$, right: residuals $\|F(c_k) - u^\dagger\|_2$. Dots indicate stopping by Morozov's discrepancy principle with $\tau = 1$ fig:noisy

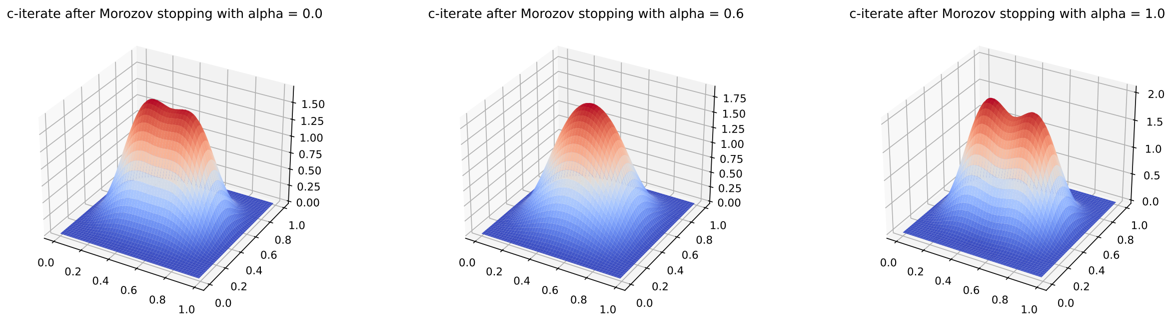

Figure 7: Results with 1% of relative noise, left to right: iterates w_k^δ stopped by Morozov's discrepancy principle with $\tau = 1$ for $\alpha = 0$ (non-accelerated Levenberg-Marquardt method), $\alpha = 0.6$ and $\alpha = 1.0$ fig:noisy

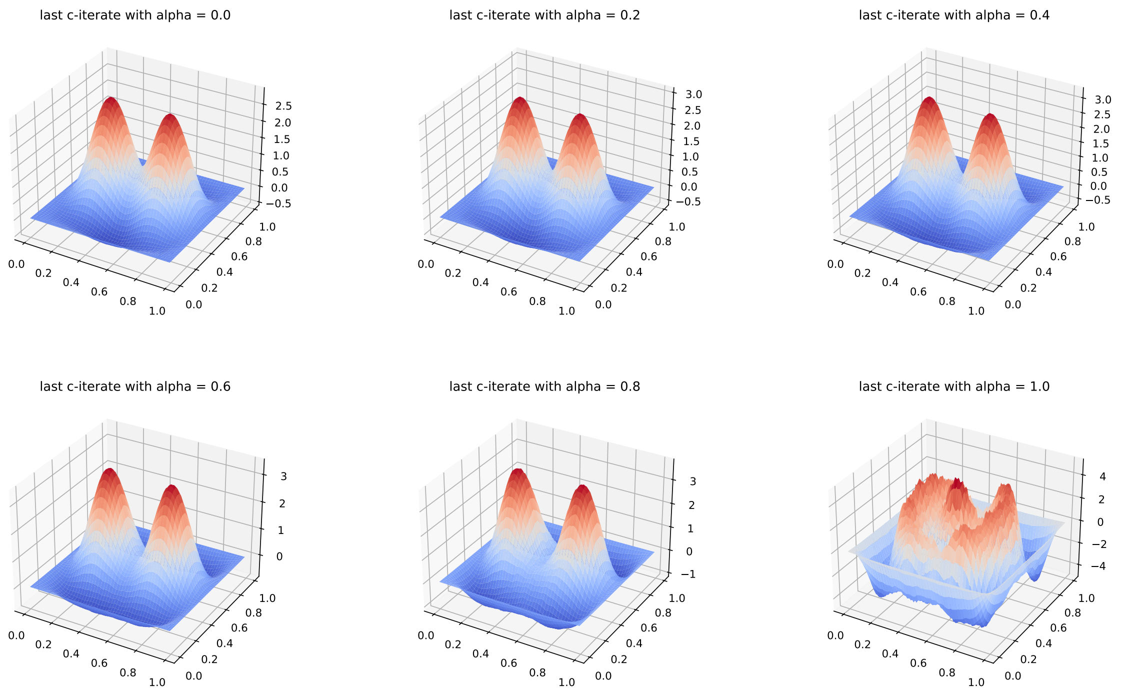

Figure 8: Results with 1% of relative noise, left to right: iterates w_{100} for $\alpha = k/10$ for $k = 0, 2, ..., 10$ fig:noisy

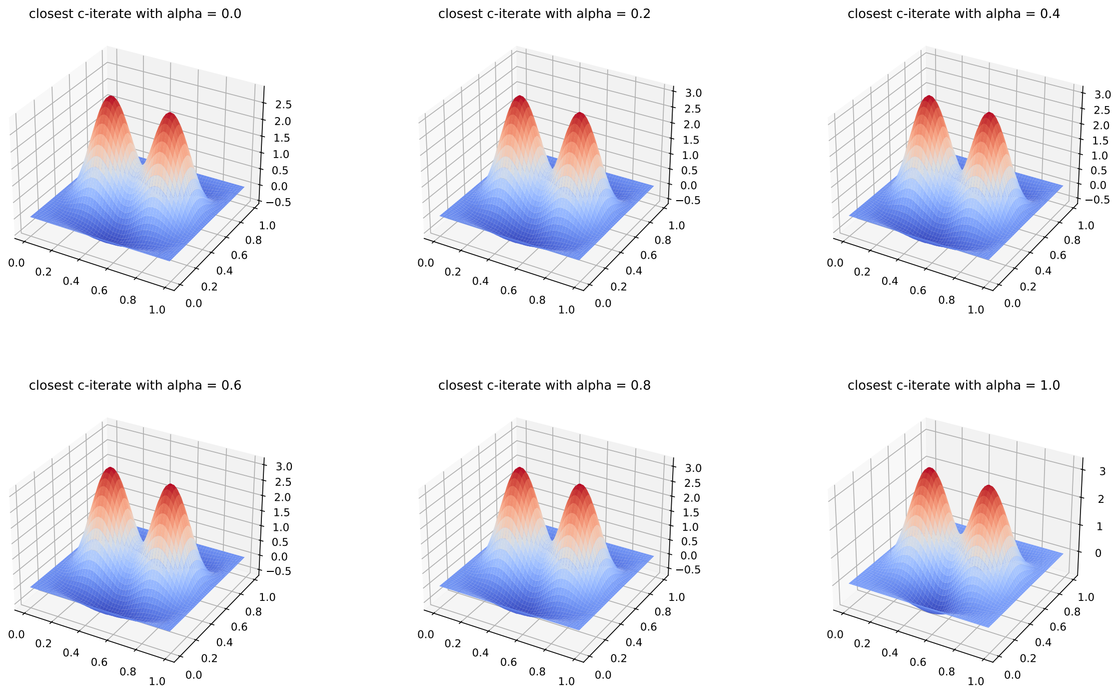

Figure 9: Results with 1% of relative noise, left to right: iterates $w_{k^*(\alpha)}$ for $\alpha = k/10$ for $k = 0, 2, ..., 10$, where $k^*(\alpha)$ is the index where w_k is closest to $c^\dagger$ in Euclidean norm fig:noisy

3.2 An inverse problem in neural network training

ssec:num-nnp

In this section, the problem of forecasting the concentration of CO in a gas sensor array is considered. Since we already used this model problem for numerical experiments in [25, 26], we are here brief in the description.

We utilize a dataset obtained from a gas delivery platform facility at the ChemoSignals Laboratory in the BioCircuits Institute at the University of California, San Diego (the actual data utilized here can be accessed on the UC Irvine Machine Learning Repository at https://archive.ics.uci.edu/ml/index.php, specifically under the dataset titled *Gas sensor array under dynamic gas mixtures*).

Formulation of the inverse problem. This dataset comprises readings from 16 different chemical sensors exposed to varying concentrations of a mixture of Ethylene and CO in the air. The measurements were obtained through continuous acquisition of signals from the 16-sensor array over approximately 12 hours without interruption; each sensor data consists of $N = 4,188,262$ scalar measurements (for a comprehensive description of the experiment, please see [11, 25]).

We address the inverse problem proposed in [25, 26] namely, to predict the reading from sensor #16, the last sensor, by leveraging the readings from the preceding sensors (see [25, Figure 3] for scatter plots of sensor $\#i$ readings against sensor #16 readings, for $i \leq 15$). As in [26], we employ a neural network (NN) in this context, which takes the readings from the first sensors as input and produces a scalar value as output, predicting the reading of the last sensor. Following [26], the structure of the NN used in our experiments reads:

— Input: $z \in \mathbb{R}^{14}$, readings of the first 14 sensors;[3]

[3]Sensor #2 readings are excluded due to significant lack of accuracy; see [25].

— Output: $NN(z; W, b) = \sigma(Wz + b) \in \mathbb{R}$.
Here $W \in R^{1\times 14}$ is a matrix of weights, $b \in R$ is a scalar bias, and $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function defined by

$$\sigma(t) = \begin{cases} c + a(t-c) & , \ t \geq c \\ t & , \ -c < t < c \\ -c + a(t+c) & , \ t \leq -c \end{cases} , \tag{28}$$

eq:sigma

where $0 < a < 1$ and $c > 0$. This is a variation of the *saturated linear activation function* [7] (the constants a and c should be chosen s.t. the range of σ contains all possible readings of sensor #16).

This is a shallow NN with only one layer (the output layer); the dimention of the corresponding parameter space is 15, the dimention of (W, b). For linear σ, this NN approach simplifies to the multiple linear regression approach considered in [25]. The inverse problem under consideration is a NN training problem, i.e. one aims to find an approximate solution to the nonlinear system

$$F_i(W, b) = y_i^\delta \,, \quad i = 0, \ldots, N_t - 1,$$

where $F_i(W, b) := NN(z_i; W, b) = \sigma(Wz_i + b)$. Here $N_t < N$ is the size of the training set and $z_i \in \mathbb{R}^{14}$ contains the readings of sensors $(\#1, \#3, \#4, \ldots, \#15)$, for $i = 0, \ldots, N_t - 1$. To suit our objectives, it is advantageous to express the preceding system in the form

$$\mathbf{F}(W, b) \ = \mathbf{y}^\delta \,, \tag{29}$$

eq:ip-nn

where $\mathbf{F}(W, b) := \big[F_i(W, b)\big]_{i=0}^{N_t-1}$ and $\mathbf{y}^\delta = \big[y_i^\delta\big]_{i=0}^{N_t-1}$.

rem:wTCC **Remark 3.1** (On the choice of the activation function)**.** *The real function $\sigma(t)$ in (28) is not differentiable at $t = -c$ and $t = c$. Consequently, the theoretical findings discussed in Section 2 cannot be applied to the inverse problem in (29) (indeed, the operator $\mathbf{F}$ does not satisfy (A1), (A2)). However, one observes that:*

Defining $s(t) := \partial_+ \sigma(t)$, the right derivative of σ at $t \in \mathbb{R}$, a direct calculation shows that

$$\|\sigma(t') - \sigma(t) - s(t)(t'-t)\| \ \leq \ \widetilde{\eta} \|\sigma(t') - \sigma(t)\| \ \ \textit{for all } t,\, t' \in \mathbb{R} \tag{30}$$

eq:tcc-s

with $\widetilde{\eta} = (1-a)a^{-1}$. Therefore, for each $0 \leq i < N_t$ the operator $F_i : (W, b) \mapsto \sigma(Wz_i + b)$, with σ as in (28), satisfies (A2) in $\mathbb{R}^{14} \times \mathbb{R}$ with $F_i'(W, b)$ replaced by $\widetilde{F}_i'(W, b) : \mathbb{R}^{14} \times \mathbb{R} \ni (W_h, b_h) \mapsto s(Wz_i + b)(W_h z_i + b_h) \in \mathbb{R}$; the corresponding constant in (A2) reads $\eta_i = (1-a)a^{-1} \max\{\|z_i\|, 1\}$. An immediate consequence of these facts is that $\mathbf{F} : (W, b) \mapsto \big[F_i(W, b)\big]_{i=0}^{N_t-1}$ in (29) satisfies (A2) in $\mathbb{R}^{14} \times \mathbb{R}$, with F_i' replaced by $\widetilde{F}'$, for $\eta = \max_i\{\eta_i\}$.

It is well known that convergence proofs of nonlinear Landweber and LM methods can be derived under assumption (A2) where F' does not necessarily have to be the derivative of F (see [15]); it only needs to be a linear operator that is uniformly bounded in a neighborhood of the initial guess x_0. We conjecture that the results obtained in Section 2 can be extended to the framework described above. This is part of our ongoing work.

For a given pair of parameters (W, b), the performance $\mathcal{P}$ of the corresponding neural network $NN(\cdot; W, b)$ is defined by

$$\mathcal{P}(NN(\cdot; W, b)) \ := \ 1 - \frac{1}{N_T} \sum_{i=N_t}^{N_t+N_T-1} \frac{\|NN(z_i; W, b) - y_i^\delta\|}{\|y_i^\delta\|} \,, \tag{31}$$

def:perf

were $N_T \in \mathbb{N}$ is the size of the test set. The sum in the above definition gives the average misfit betwen the predicted value $NN(z_i; W, b)$ and y_i^δ, evaluated over the test set $\{z_i,\ N_t \leq i < N_t + N_T - 1\}$. Notice that $0 \leq \mathcal{P}(NN(\cdot; W, b)) \leq 1$ for all (W, b), while $\mathcal{P}(NN(\cdot; W, b)) = 1$ is the best possible performance.

Remark 3.2 (The training set and test set)**.** *The 'training set' and 'test set' are comprised of samples with sizes of N_t and N_T respectively. In our numerical experiments we use $N_t = 4,000,000$ and $N_T = 100,000$ (notice that $Nt + N_T < N$).*

Numerical implementations. In what follows the inLM method is implemented for solving the NN training problem (29). In view of Remark 3.1 we choose $a = \frac{2}{3}$ and $c = 8$ in (28). Consequently, we generate an activation function $\bar{\sigma}$ of the form (28), satisfying (30) for $\eta = 0.5$.

The sensor readings $(z_i, y_i^\delta) \in \mathbb{R}^{14} \times \mathbb{R}$ on the training set are scaled by the factor $\max_{i \le N_t} \|z_i\|$. An analogous procedure is performed on the test set. Consequently, after scaling, it holds $\|z_i\| \le 1$, for $i = 0, \dots, N_t + N_T$. From Remark 3.1 it follows that, for $\bar{\sigma}$ as above, all operators $F_i(W, b)$ satisfy (A2) (with F_i' replaced by $\widetilde{F}_i'$) for the same constant $\eta = 0.5$; the same holds for the operator $\mathbf{F}$ in (29).

In our experiments the initial guess (W_0, b_0) is a random vector with coordinate values ranging in $(-1, 1)$. We approximate the linear solve in step [2.1] of Algorithm 2 by three steps of the conjugate gradient method with zero initial value. Three different runs of the inLM method are presented, each one for a different choice of (constant) inertial parameter α_k, namely $\{0.05, 0, 10, 0.20\}$.

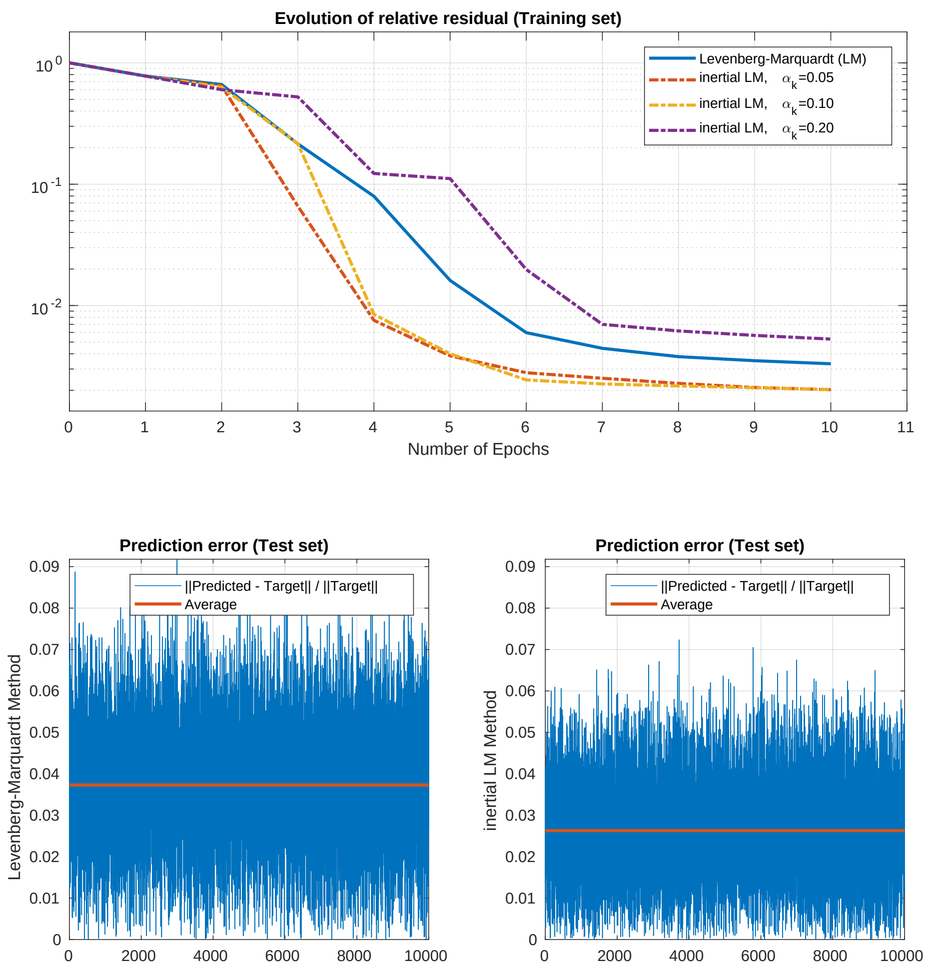

Figure 10: Neural Network training problem. (TOP) Evolution of relative residual for different methods; (BOTTOM) Prediction error of the trained NN for the test-set: LM method (left) and inLM method (right). fig:inert

For comparison purposes the classical LM method ($\alpha_k = 0$) was also implemented. Since the noise level δ is not known, all methods are computed for ten steps;[4] after the tenth step the residual evolution stagnates for all methods. The obtained results are summarized in Figure 10:

- (TOP) Evolution of relative residual $\sum_{i=0}^{N_t-1} \frac{\|NN(z_i;W_k,b_k)-y_i^\delta\|}{\|NN(z_i;W_0,b_0)-y_i^\delta\|}$ on the training set – all methods.
- (BOTTOM-RIGHT) inLM method: relative prediction error $\|NN(z_i;W_{k^*},b_{k^*})-y_i^\delta\|/\|y_i^\delta\|$ is plotted for the test set $\{z_i,\ N_t \le i < N_t+N_T-1\}$ (BLUE); the average value (RED) is 0.026. The performance of the trained Neural Network amounts to **97**%.
- (BOTTOM-LEFT): For comparison, the prediction accuracy of the NN trained by the LM method is plotted for the same test set (BLUE), the average value is 0.037 (RED). The performance of the trained Neural Network amounts to **96**%.

Here are a few observations from our numerical experiments:

- For constant choices of α_k, small values yield the best results (in our experiments $\alpha_k =$ 0.05 and 0.10). For even smaller constant values, such as $\alpha_k = 0.01$, the performance of the inLM method becomes very similar to that of the LM method (which corresponds to $\alpha_k = 0$).
- For larger constant values of α_k, e.g. $\alpha_k = 0.20$, the inLM method becomes unstable and its performance deteriorates compared to that of the LM method.
- The Neural Network trained using the inLM method outperforms the one trained with the LM method. Additionally, the inLM method converges faster. The residual decay for the inLM method stagnates after 6 steps, whereas it takes 10 steps for the LM method (see Figure 10).

4 Final remarks and conclusions

sec:conclusions

In this manuscript we propose and analyze an implicit inertial type iteration, namely the *inertial Levenberg Marquardt* (inLM) method, as an alternative for obtaining stable approximate solutions to nonlinear ill-posed operator equations. This new method can be considered as an extension of the classical Levenberg Marquardt (LM) method (indeed, if the inertial parameters α_k are set to zero the inLM reduces to the LM method).

The main results discussed in this notes are: boundedness of the sequences (x_k) and (w_k) generated by the inLM method (Propositions 2.6 and 2.7), strong convergence for exact data (Theorem 2.9), stability and semi-convergence for noisy data (Theorems 2.14 and 2.15 respectively). We also provide a bound for the stopping index in the noisy data case (Proposition 2.13).

In Section 3 two distinct ill-posed problems are used to investigate the numerical efficiency of the proposed inLM method: A parameter identification problem in an elliptic PDE and an inverse problem in neural network training.

The preliminary results obtained in our numerical experiments indicate a better performance (faster convergence) of the inLM method when compared to the LM method. The inLM method not only converges faster than the LM method (as shown in Figures 6 and 10), but it also attains an approximate solution with a significantly smaller residual in the second inverse problem.

[4] Each step corresponds to an epoch.

Acknowledgments

AL acknowledges support from the AvH Foundation. Significant part of this manuscript was writen while this author was on sabbatical leave at EMAp, Getulio Vargas Fundation, Rio de Janeiro, Brazil. DAL acknowledges support from the AvH foundation.

References

AA01 [1] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Anal. **9** (2001), no. 1-2, 3–11.

AABR02 [2] F. Alvarez, H. Attouch, J. Bolte, and P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl. (9) **81** (2002), no. 8, 747–779. MR 1930878

APR16 [3] Hedy Attouch, Juan Peypouquet, and Patrick Redont, *Fast convex optimization via inertial dynamics with Hessian driven damping*, J. Differential Equations **261** (2016), no. 10, 5734–5783. MR 3548269

Bau87 [4] J. Baumeister, *Stable Solution of Inverse Problems*, Advanced Lectures in Mathematics, Friedr. Vieweg & Sohn, Braunschweig, 1987. MR 889048

BeTe09 [5] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. **2** (2009), no. 1, 183–202.

BLS20 [6] R. Boiger, A. Leitão, and B.F. Svaiter, *Range-relaxed criteria for choosing the Lagrange multipliers in nonstationary iterated Tikhonov method*, IMA Journal of Numerical Analysis **40** (2020), no. 1, 606–627.

CoPe20 [7] P.L. Combettes and J.-C. Pesquet, *Deep neural network structures solving variational inequalities*, Set-Valued Var. Anal. **28** (2020), no. 3, 491–518.

CBL11 [8] A. De Cezaro, J. Baumeister, and A. Leitão, *Modified iterated Tikhonov methods for solving systems of nonlinear ill-posed equations*, Inverse Probl. Imaging **5** (2011), no. 1, 1–17.

Eng87 [9] H.W. Engl, *On the choice of the regularization parameter for iterated Tikhonov regularization of ill-posed problems*, J. Approx. Theory **49** (1987), no. 1, 55–63.

EngHanNeu96 [10] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.

FSHM15 [11] J. Fonollosa, S. Sheik, R. Huerta, and S. Marco, *Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring*, Sensors and Actuators B: Chemical **215** (2015), 618–629.

GS00 [12] C. W. Groetsch and O. Scherzer, *Non-stationary iterated Tikhonov-Morozov method and third-order differential equations for the evaluation of unbounded operators*, Math. Methods Appl. Sci. **23** (2000), no. 15, 1287–1300.

HG98 [13] M. Hanke and C. W. Groetsch, *Nonstationary Iterated Tikhonov Regularization*, J. Optim. Theory Appl. **98** (1998), no. 1, 37–53.

HanNeuSch95 [14] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of Landweber iteration for nonlinear ill-posed problems*, Numer. Math. **72** (1995), 21–37.

KalNeuSch08 [15] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, Radon Series on Computational and Applied Mathematics, vol. 6, Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

KN08 [16] S. Kindermann and A. Neubauer, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging **2** (2008), no. 2, 291–299.

Kir96 [17] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Applied Mathematical Sciences, vol. 120, Springer-Verlag, New York, 1996.

LP15 [18] Dirk A. Lorenz and Thomas Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vision **51** (2015), no. 2, 311–325. MR 3314536

Mar70 [19] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle **4** (1970), no. Sér. R-3, 154–158.

MO03 [20] A. Moudafi and M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, J. Comput. Appl. Math. **155** (2003), no. 2, 447–454.

Nes83 [21] Y.E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR **269** (1983), no. 3, 543–547.

Nes04 [22] Yurii Nesterov, *Introductory lectures on convex optimization*, Applied Optimization, vol. 87, Kluwer Academic Publishers, Boston, MA, 2004, A basic course. MR 2142598

Pol64 [23] B. T. Poljak, *Some methods of speeding up the convergence of iterative methods*, Ž. Vyčisl. Mat i Mat. Fiz. **4** (1964), 791–803. MR 169403

RLM24 [24] J. Rabelo, A. Leitão, and A.L. Madureira, *On inertial iterated-tikhonov methods for solving linear ill-posed problems*, Inverse Problems **40** (2024), no. 3, 035002.

RSL22 [25] J.C. Rabelo, Y. Saporito, and A. Leitão, *On stochastic Kaczmarz type methods for solving large scale systems of ill-posed equations*, Inverse Problems **38** (2022), no. 2, 025003.

RSL24 [26] ______, *On projective stochastic-gradient type methods for solving large scale systems of nonlinear ill-posed equations: Applications to machine learning*, Inverse Problems (2024), submitted.

Roc76 [27] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976), no. 5, 877–898.

Sch93a [28] O. Scherzer, *Convergence rates of iterated Tikhonov regularized solutions of nonlinear ill-posed problems*, Numer. Math. **66** (1993), no. 2, 259–279.

SBC16 [29] Weijie Su, Stephen Boyd, and Emmanuel J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, J. Mach. Learn. Res. **17** (2016), Paper No. 153, 43. MR 3555044