

# A data-driven iteratively regularized Landweber iteration

A. Aspri<sup>2</sup>, S. Banert<sup>3</sup>, O. Öktem<sup>3</sup>, and O. Scherzer<sup>1,2</sup>

<sup>1</sup>Computational Science Center, University of Vienna, [otmar.scherzer@univie.ac.at](mailto:otmar.scherzer@univie.ac.at)

<sup>2</sup>Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz,  
[andrea.aspri@ricam.oeaw.ac.at](mailto:andrea.aspri@ricam.oeaw.ac.at)

<sup>3</sup>KTH Royal Institute of Technology, Stockholm, [{banert,ozan}@kth.se](mailto:{banert,ozan}@kth.se)

March 6, 2020

## Abstract

We derive and analyse a new variant of the iteratively regularized Landweber iteration, for solving linear and nonlinear ill-posed inverse problems. The method takes into account training data, which are used to estimate the interior of a black box, which is used to define the iteration process. We prove convergence and stability for the scheme in infinite dimensional Hilbert spaces. These theoretical results are complemented by some numerical experiments for solving linear inverse problems for the Radon transform and a nonlinear inverse problem for Schlieren tomography.

**Keywords.** Iteratively regularized Landweber iteration, expert and data driven regularization, black box strategy.

2010 AMS subject classifications. 65J20 (65J10, 65J15, 65J22)

## 1. INTRODUCTION

This paper is concerned with a generalization of the *iteratively regularized Landweber iteration*, as introduced in [17], for solving (linear and nonlinear) ill-posed operator equations

$$F(u) = y. \quad (1.1)$$

In the course of this paper, for the sake of simplicity of presentation, we restrict our attention to an operator  $F : \mathcal{D}(F) \subset X \rightarrow Y$  between real separable Hilbert spaces  $X$  and  $Y$  with inner products  $\langle \cdot, \cdot \rangle$  and norms  $\| \cdot \|$ , respectively. We denote with  $y^\delta$  noisy data and we assume that

$$\|y^\delta - y\| \leq \delta.$$

Generalizations to the Banach space setting (see [19]) are formally similar, but technically more complicated, and thus omitted here.

In the original form, the iteratively regularized Landweber iteration, as considered in [17] (see also [9]), consists in computing the following iterative updates

$$u_{k+1} := u_k - F'(u_k)^*(F(u_k) - y^\delta) - \lambda_k(u_k - u^{(0)}), \quad k \in \mathbb{N}, \quad (1.2)$$

where  $u^{(0)}$  is an initial guess which incorporates *a-priori knowledge* on the solution to be recovered. In presence of noise in the data (that is the available data are  $y^\delta$ ), to guarantee that the iterative scheme (1.2) is a regularization procedure, it has to be complemented with a stopping rule. The discrepancy principle is often employed, i.e., the iteration is stopped after the first  $k_* = k_*(\delta, y^\delta)$  steps for which:

$$\|F(u_{k_*}) - y^\delta\| \leq \tau\delta < \|F(u_k) - y^\delta\|, \quad 0 \leq k < k_*,$$

for some  $\tau > 1$ . Method (1.2) can be considered a modification of the Landweber iteration, i.e.,

$$u_{k+1} := u_k - F'(u_k)^*(F(u_k) - y^\delta), \quad k \in \mathbb{N}, \quad (1.3)$$

when we put  $\lambda_k = 0$  for all  $k \in \mathbb{N}$ . However, opposed to the Landweber iteration (1.3), the convergence rates analysis of (1.2) is indeed simpler, and requires less restrictive conditions on the operator  $F$  (see [9]).

The reason for that is the presence of the damping term  $\lambda_k(u_k - u^{(0)})$  which stabilizes the algorithm: The modified Landweber iteration converges to a solution which is closest to  $u^{(0)}$ , which is not guaranteed for the Landweber iteration itself without posing additional assumptions. However, the practical convergence rates are even slower than for the Landweber iteration. The damping term  $\lambda_k(u_k - u^{(0)})$  was originally introduced as an additional stabilizing term in the Gauß-Newton's method in [3]. Later, it has been observed that it has a similar effect in all the iterative regularization methods when used as an add-on factor. This observation is the motivation for this paper to introduce a data driven damping factor in the Landweber iteration. We first note that the iteration (1.2) can be rewritten in the form

$$u_{k+1} := u_k - \frac{1}{2} \partial_u \left( \|F(u) - y^\delta\|^2 + \lambda_k \|u - u^{(0)}\|^2 \right) (u_k), \quad k \in \mathbb{N}. \quad (1.4)$$

Our objective is to include a damping factor gained from *expert data*  $(u^{(i)}, F(u^{(i)}))_{1 \leq i \leq n}$  in the iteration process. As a first attempt, one could generalize (1.4) by ignoring the image data  $(F(u^{(i)}))_{1 \leq i \leq n}$ , which leads to

$$u_{k+1} := u_k - \frac{1}{2} \partial_u \left( \|F(u) - y^\delta\|^2 + \lambda_k \sum_{i=1}^n \|u - u^{(i)}\|^2 \right) (u_k), \quad k \in \mathbb{N}. \quad (1.5)$$

From an analytical point of view, this does not offer significant benefits because we expect that in general it will converge to the solution which is closest to the mean  $\hat{u} = \sum_{i=1}^n u^{(i)}$ . Therefore, in order to include the image data  $(F(u^{(i)}))_{1 \leq i \leq n}$  as prior information as well, we follow a *black box strategy*. To be more specific, we identify an operator  $\hat{A}$ , which maps each  $u^{(i)}$  to  $F(u^{(i)})$ ,  $i = 1, \dots, n$ , and vice versa, to include as a damping term in the Landweber iteration (1.3), i.e.,

$$u_{k+1} := u_k - \frac{1}{2} \partial_u \left( \|F(u) - y^\delta\|^2 + \lambda_k^\delta \|\hat{A}(u) - y^\delta\|^2 \right) (u_k), \quad k \in \mathbb{N}, \quad (1.6)$$

which, in explicit form, is equal to

$$u_{k+1} := u_k - F'(u_k)^* (F(u_k) - y^\delta) - \lambda_k^\delta \hat{A}'(u_k)^* (\hat{A}(u_k) - y^\delta), \quad k \in \mathbb{N}. \quad (1.7)$$

The objective of the second term in (1.7) is to give some bias for the expert data. We mention that *system identification* and *black box* theory was considered extensively in the sixties of the last century, see for instance [15]. Our main results concern the proof of strong convergence and stability for scheme (1.7), essentially under the assumptions that  $F$  satisfies the usual tangential cone condition, see (2.3), and both  $F$  and  $\hat{A}$  are Fréchet differentiable. These theoretical results are followed by some numerical experiments for linear and nonlinear operator equations (1.1). To be more precise, we take the Radon (see [12]) and the Schlieren (see [18]) operators as models of linear and nonlinear problems, respectively. Concerning the operator  $\hat{A}$ , we do not tackle the problem in the full generality in the sense that all the numerical experiments are realized by restricting attention to bounded linear operators between separable Hilbert spaces  $X$  and  $Y$ , that is, taking  $\tilde{A} \in \mathcal{B}(X, Y)$ , we consider

$$u_{k+1} := u_k - F'(u_k)^* (F(u_k) - y^\delta) - \lambda_k^\delta \tilde{A}^* (\tilde{A}u_k - y^\delta), \quad k \in \mathbb{N}. \quad (1.8)$$

We reserve to point the general case of nonlinear operators for future work. Specifically, given  $n$  pairs of input-output relation  $(u^{(i)}, F(u^{(i)}))$ , and defining the functional

$$l(\tilde{T}) := \frac{1}{2} \sum_{i=1}^n \|\tilde{T}u^{(i)} - y^{(i)}\|_Y^2,$$

where  $\tilde{T} \in \mathcal{B}(X, Y)$  and  $y^{(i)} = F(u^{(i)})$ , we define  $\tilde{A} : X \rightarrow Y$  as the bounded linear operator which satisfies

$$l(\tilde{A}) = \min_{\tilde{T} \in \mathcal{B}(X, Y)} l(\tilde{T}) = \min_{\tilde{T} \in \mathcal{B}(X, Y)} \left[ \frac{1}{2} \sum_{i=1}^n \|\tilde{T}u^{(i)} - y^{(i)}\|_Y^2 \right]. \quad (1.9)$$

The operator  $\tilde{A}$  can be projected onto a finite dimensional subspace and here represented by a matrix  $A \in \mathbb{R}^{M \times N}$ , where  $M, N \in \mathbb{N}$  will be defined in Section 3. In fact, from classical results of functional analysis, every bounded linear operator has a matrix representation of infinite dimensions, through complete orthonormal bases in  $X$  and  $Y$ , see for example [10]. We refer the reader to Section 3 for a more in-depth discussion. In all the numerical experiments we compare the outcomes of the data

driven iteratively regularized Landweber scheme (1.8) with the classic iteratively regularized Landweber scheme (1.2) and the Landweber iteration (1.3). This is a first attempt to include expert data in iterative reconstruction algorithms. Nonlinear learning strategies might behave significantly better. However, for the solution of linear ill-posed operator equations, the numerical reconstructions of (1.8) are significantly better than the results of Landweber iterations (1.2) and (1.3). In the case of nonlinear operator equations this approach is not as well suited, because we attempt to simulate a nonlinear input-output relation by a linear operator  $\hat{A}$ . This is a challenging topic to investigate for future research.

The outline of the paper is as follows: In Section 2 we analyze (1.7) in an infinite dimensional Hilbert space setting and prove strong convergence and stability. In Section 3 we discuss the numerical implementation of (1.8) and study applications to the Radon inversion, for the linear case, and to the Schlieren tomography, for the nonlinear case, with full and limited data. In all the numerical examples we compare the outcomes of (1.2), (1.3) and (1.8).

## 2. ITERATIVELY REGULARIZED LANDWEBER ITERATION

In this section we analyze the convergence of the iteratively regularized Landweber iteration, presented in the introduction, which we recall here for the reader's sake

$$\begin{aligned} u_{k+1} &:= u_k - F'(u_k)^*(F(u_k) - y^\delta) - \lambda_k^\delta \hat{A}'(u_k)^*(\hat{A}(u_k) - y^\delta), \quad k \in \mathbb{N}, \\ u_0 &\text{ - initial guess,} \end{aligned} \quad (2.1)$$

where  $y^\delta$  are noisy data such that

$$\|y^\delta - y\| \leq \delta,$$

and  $\lambda_k^\delta$  is a suitable parameter. Moreover, we assume that (1.1) has a solution  $u^\dagger$ , possibly not unique.

For the analysis, we basically follow the approach proposed in [9, Chapters 1, 3] because such iterates can be considered a modified version of the Landweber scheme.

It is well-known that for iterative schemes and in case of noisy data, a stopping criterion has to be added to the scheme (2.1) in order to obtain a regularization method. For that, we employ the discrepancy principle for which the iteration is stopped after  $k_* = k_*(\delta, y^\delta)$  steps once

$$\|F(u_{k_*}) - y^\delta\| \leq \tau\delta \leq \|F(u_k) - y^\delta\|, \quad 0 \leq k < k_*, \quad (2.2)$$

where  $\tau$  is a positive number chosen suitably.

First, we introduce some assumptions on the operators  $F$  and  $\hat{A}$  to guarantee local convergence of the iterative schemes as (2.1), and we define some constants which will appear in the convergence theorem.

**Assumption 2.1** Let  $\mathcal{B}_\rho(u^\dagger) \subseteq \mathcal{B}_{2\rho}(u_0)$ , where  $\mathcal{B}_\rho(u^\dagger)$  denotes a closed ball of radius  $\rho$  and center  $u^\dagger$ . We assume that

(i)  $F$  satisfies the *tangential cone condition*, that is

$$\|F(u) - F(v) - F'(u)(u - v)\| \leq \nu \|F(u) - F(v)\|, \quad \forall u, v \in \mathcal{B}_\rho(u^\dagger), \quad (2.3)$$

where  $\nu > 0$ .

(ii)  $F$  and  $\hat{A}$  have continuous Fréchet derivative  $F'$  and  $\hat{A}'$ , respectively, with Lipschitz constants  $L_F$  and  $L_{\hat{A}}$ , i.e., for all  $u \in \mathcal{B}_\rho(u^\dagger)$  it holds

$$\|F'(u)\| \leq L_F, \quad (2.4a)$$

$$\|\hat{A}'(u)\| \leq L_{\hat{A}}. \quad (2.4b)$$

(iii) We assume that data-driven model  $\hat{A}$  cannot fully explain the model for the true data, hence

$$\|\hat{A}(u^\dagger) - y\| \geq C_N \text{ with } C_N > 0. \quad (2.5)$$

Here, we present an auxiliary estimate which will be needed to prove a monotonicity property and some convergence results of (2.1).

**Lemma 2.2** *Let Assumption 2.1 be satisfied. Then there exists a positive constant  $C_{\hat{A}}^\delta$  such that*

$$\left\| \hat{A}(u_k) - y^\delta \right\| \leq C_{\hat{A}}^\delta \quad (2.6)$$

holds true for every  $u_k \in \mathcal{B}_\rho(u^\dagger)$ .

*Proof:* Equation (2.6) follows from the triangle inequality and the Lipschitz-continuity of  $\hat{A}'$ , in fact

$$\begin{aligned} \left\| \hat{A}(u_k) - y^\delta \right\| &\leq \left\| \hat{A}(u_k) - \hat{A}(u^\dagger) \right\| + \left\| \hat{A}(u^\dagger) \right\| + \|y^\delta\| \\ &= \left\| \int_0^1 \hat{A}'(u^\dagger + t(u_k - u^\dagger))(u_k - u^\dagger) dt \right\| + \left\| \hat{A}(u^\dagger) \right\| + \|y^\delta\| \\ &\leq L_{\hat{A}}\rho + \left\| \hat{A}(u^\dagger) \right\| + \|y^\delta\| =: C_{\hat{A}}^\delta. \end{aligned} \quad \square$$

We first prove that, under some suitable assumptions on  $\lambda_k^\delta$ , a monotonicity property is verified for the scheme (2.1).

**Proposition 2.3** *Let Assumption 2.1 be satisfied. Assume that  $u_k \in \mathcal{B}_\rho(u^\dagger)$  and*

$$\left\| F(u_k) - y^\delta \right\| \geq \tau\delta, \quad (2.7)$$

where  $\tau > 0$  and it is such that

$$C_\tau := 1 - L_F^2 - \nu - \frac{1 + \nu}{\tau} \geq 0. \quad (2.8)$$

Moreover, let us assume that there exists a positive constant  $C_\lambda^\delta$  such that the following conditions hold:

$$\lambda_k^\delta L_{\hat{A}} C_{\hat{A}}^\delta \leq \rho, \quad (2.9a)$$

$$\lambda_k^\delta \leq C_\lambda^\delta \left\| F(u_k) - y^\delta \right\|^2, \quad (2.9b)$$

$$C_\tau - 2L_{\hat{A}} C_{\hat{A}}^\delta C_\lambda^\delta \rho > 0. \quad (2.10)$$

Then  $u_{k+1} \in \mathcal{B}_\rho(u^\dagger)$  and  $\|u_{k+1} - u^\dagger\| \leq \|u_k - u^\dagger\|$ .

*Proof:* Let us assume that  $u_k \in \mathcal{B}_\rho(u^\dagger)$ , then we have

$$\begin{aligned} &\left\| u_{k+1} - u^\dagger \right\|^2 - \left\| u_k - u^\dagger \right\|^2 \\ &\leq 2 \left( - \langle F(u_k) - y^\delta, F'(u_k)(u_k - u^\dagger) \rangle + \left\| F'(u_k)^*(F(u_k) - y^\delta) \right\|^2 \right. \\ &\quad \left. - \lambda_k^\delta \left\langle \hat{A}(u_k) - y^\delta, \hat{A}'(u_k)(u_k - u^\dagger) \right\rangle + (\lambda_k^\delta)^2 \left\| \hat{A}'(u_k)^*(\hat{A}(u_k) - y^\delta) \right\|^2 \right) \\ &\leq 2 \left( - \left\| F(u_k) - y^\delta \right\|^2 + \langle F(u_k) - y^\delta, F(u_k) - y - F'(u_k)(u_k - u^\dagger) \rangle \right. \\ &\quad \left. + \delta \left\| F(u_k) - y^\delta \right\| + \left\| F'(u_k)^*(F(u_k) - y^\delta) \right\|^2 \right. \\ &\quad \left. - \lambda_k^\delta \left\langle \hat{A}(u_k) - y^\delta, \hat{A}'(u_k)(u_k - u^\dagger) \right\rangle + (\lambda_k^\delta)^2 \left\| \hat{A}'(u_k)^*(\hat{A}(u_k) - y^\delta) \right\|^2 \right) =: 2(T_F + T_{\hat{A}}), \end{aligned} \quad (2.11)$$

where  $T_F$  and  $T_{\hat{A}}$  represent the terms in  $F$  and  $\hat{A}$ , respectively, in the right hand side of the second inequality. Next, we study  $T_F$  and  $T_{\hat{A}}$  separately, providing for both of them a bound in terms of the square norm of the residual of  $F$  at the  $k$ -iteration. By using (2.3), (2.4a), (2.7) and (2.8) we get

$$\begin{aligned} T_F &= - \left\| F(u_k) - y^\delta \right\|^2 + \langle F(u_k) - y^\delta, F(u_k) - y - F'(u_k)(u_k - u^\dagger) \rangle \\ &\quad + \delta \left\| F(u_k) - y^\delta \right\| + \left\| F'(u_k)^*(F(u_k) - y^\delta) \right\|^2 \\ &\leq - \left\| F(u_k) - y^\delta \right\|^2 + \nu \langle F(u_k) - y^\delta, F(u_k) - y \rangle \\ &\quad + \delta \left\| F(u_k) - y^\delta \right\| + L_F^2 \left\| F(u_k) - y^\delta \right\|^2 \\ &\leq \left\| F(u_k) - y^\delta \right\| \left( -(1 - L_F^2 - \nu) \left\| F(u_k) - y^\delta \right\| + (1 + \nu) \delta \right) \\ &\leq - C_\tau \left\| F(u_k) - y^\delta \right\|^2. \end{aligned} \quad (2.12)$$

On the contrary, to estimate the term  $T_{\hat{A}}$ , we utilize (2.4b), (2.6), (2.9a) and (2.9b) from what it follows

$$\begin{aligned} T_{\hat{A}} &= -\lambda_k^\delta \left\langle \hat{A}(u_k) - y^\delta, \hat{A}'(u_k)(u_k - u^\dagger) \right\rangle + (\lambda_k^\delta)^2 \left\| \hat{A}'(u_k)^* (\hat{A}(u_k) - y^\delta) \right\|^2 \\ &\leq \lambda_k^\delta L_{\hat{A}} \left\| \hat{A}(u_k) - y^\delta \right\| \left( \rho + \lambda_k^\delta L_{\hat{A}} \left\| \hat{A}(u_k) - y^\delta \right\| \right) \\ &\leq 2\lambda_k^\delta L_{\hat{A}} C_{\hat{A}}^\delta \rho \leq 2L_{\hat{A}} C_{\hat{A}}^\delta C_{\lambda\rho}^\delta \left\| F(u_k) - y^\delta \right\|^2. \end{aligned} \quad (2.13)$$

Finally, by inserting (2.13) and (2.12) into (2.11), we get

$$\left\| u_{k+1} - u^\dagger \right\|^2 - \left\| u_k - u^\dagger \right\|^2 \leq -2(C_\tau - 2L_{\hat{A}} C_{\hat{A}}^\delta C_{\lambda\rho}^\delta) \left\| F(u_k) - y^\delta \right\|^2 \quad (2.14)$$

which gives the assertion of the theorem thanks to (2.10).  $\square$

**Remark 2.1** We note that (2.9a) is asymptotically always satisfied if the iteration is convergent and the damping parameters (2.9) are chosen sufficiently small. Equation 2.10 requires that the operator  $F$  is not too non-linear and that the iterates are close to the true solution.

To prove a convergence result for the iteratively regularized Landweber iteration (2.1), we first need to show that there exists a finite stopping index  $k_*$  for which the discrepancy principle (2.2) holds.

**Corollary 2.4** Under the assumptions of Proposition 2.3, let  $k_*$  be the first index such that the discrepancy principle (2.2) is verified. Then

$$k_*(\tau\delta)^2 \leq \sum_{k=0}^{k_*-1} \left\| F(u_k) - y^\delta \right\|^2 \leq \frac{1}{2(C_\tau - 2L_{\hat{A}} C_{\hat{A}}^\delta C_{\lambda\rho}^\delta)} \left\| u_0 - u^\dagger \right\|. \quad (2.15)$$

In particular, when  $\delta = 0$ , we have

$$\sum_{k=0}^{\infty} \left\| F(u_k) - y \right\|^2 < \infty. \quad (2.16)$$

*Proof:* We sum the inequality (2.14) from 0 to the step  $k_* - 1$ , that is

$$\sum_{k=0}^{k_*-1} \left[ \left\| u_{k+1} - u^\dagger \right\|^2 - \left\| u_k - u^\dagger \right\|^2 \right] \leq -2(C_\tau - 2L_{\hat{A}} C_{\hat{A}}^\delta C_{\lambda\rho}^\delta) \sum_{k=0}^{k_*-1} \left\| F(u_k) - y^\delta \right\|^2,$$

hence

$$\left\| u_0 - u^\dagger \right\|^2 - \left\| u_{k_*} - u^\dagger \right\|^2 \geq 2(C_\tau - 2L_{\hat{A}} C_{\hat{A}}^\delta C_{\lambda\rho}^\delta) \sum_{k=0}^{k_*-1} \left\| F(u_k) - y^\delta \right\|^2. \quad (2.17)$$

Using the fact that  $\left\| u_0 - u^\dagger \right\|^2 - \left\| u_{k_*} - u^\dagger \right\|^2 \leq \left\| u_0 - u^\dagger \right\|^2$  and the discrepancy principle (2.2), from (2.17) it is straightforward to obtain (2.15). Moreover, when  $\delta = 0$ , we can assume  $\tau$  arbitrarily large such that  $C_\tau - 2L_{\hat{A}} C_{\hat{A}}^0 C_{\lambda\rho}^0 \xrightarrow{\tau \rightarrow \infty} 1 - \nu - L_F^2 - 2L_{\hat{A}} C_{\hat{A}}^0 C_{\lambda\rho}^0$  and  $k_* \rightarrow \infty$ . Therefore (2.15) reduces to

$$\sum_{k=0}^{\infty} \left\| F(u_k) - y \right\|^2 \leq \frac{1}{2(1 - \nu - L_F^2 - 2L_{\hat{A}} C_{\hat{A}}^0 C_{\lambda\rho}^0)} \left\| u_0 - u^\dagger \right\|,$$

which gives the assertion of the theorem.  $\square$

Next, before proving the convergence result in case of noisy data, we show that, when  $\delta = 0$ , the residual norm of the modified Landweber iteration (2.1) goes to zero as  $k$  tends to infinity. This means that if the iterations converge, then the limit is a solution of (1.1).

**Theorem 2.5** Let Assumption 2.1, (2.6) and (2.9b) hold, with  $\delta = 0$ . Then the iteratively regularized Landweber iteration (2.1), applied to the exact data  $y$ , converges to a solution  $u^\dagger$  of  $F(u) = y$ .

*Proof:* Let us define

$$e_k = u_k - u^\dagger. \quad (2.18)$$

Given  $j \geq k$ , we choose an integer  $h$  such that  $k \leq h \leq j$  and

$$\|F(u_h) - y\| \leq \|F(u_i) - y\|, \quad \forall k \leq i \leq j. \quad (2.19)$$

Then, we consider

$$\|e_j - e_k\| \leq \|e_j - e_h\| + \|e_h - e_k\|. \quad (2.20)$$

We only analyse the first term in (2.20) since the other one can be treated in the same way. So, we have

$$\|e_j - e_h\|^2 = \langle e_j - e_h, e_j - e_h \rangle = 2\langle e_h - e_j, e_h \rangle + \|e_j\|^2 - \|e_h\|^2. \quad (2.21)$$

We only need to study the first term in the right-hand side of the previous equality since  $\|e_j\|^2$  and  $\|e_h\|^2$  are monotonically decreasing to some  $\varepsilon \geq 0$ , thanks to Proposition 2.3 (even in the case  $\delta = 0$ ), hence  $\|e_j\|^2 - \|e_h\|^2$  converges to 0 as  $k \rightarrow \infty$ . We show that  $|\langle e_h - e_j, e_h \rangle| \rightarrow 0$  as  $k \rightarrow +\infty$ . Using the definition of  $e_j, e_h$ , see (2.18), the iteration (2.1) and the fact that  $j \geq h$ , we find

$$\begin{aligned} |\langle e_h - e_j, e_h \rangle| &= |\langle u_j - u_h, e_h \rangle| \\ &\leq \left| \left\langle \sum_{i=h}^{j-1} F'(u_i)^*(y - F(u_i)) + \lambda_i \hat{A}'(u_i)^*(y - \hat{A}(u_i)), e_h \right\rangle \right| \\ &\leq \sum_{i=h}^{j-1} \left| \left\langle y - F(u_i), F'(u_i)(u_h - u^\dagger) \right\rangle \right| + \sum_{i=h}^{j-1} \lambda_i \left| \left\langle y - \hat{A}(u_i), \hat{A}'(u_i)(u_h - u^\dagger) \right\rangle \right| \\ &=: S_F + S_{\hat{A}} \end{aligned} \quad (2.22)$$

We study the two terms  $S_F$  and  $S_{\hat{A}}$  separately.

*Term  $S_F$ :* the analysis can be found, for example, in [9]. For the sake of completeness, here we summarize the principal steps to obtain an estimate of  $S_F$ . Adding and subtracting  $u_i$  and  $y - F(u_i)$  in  $S_F$ , we find

$$S_F \leq \sum_{i=h}^{j-1} [\|y - F(u_i)\| (\|y - F(u_i) - F'(u_i)(u^\dagger - u_i)\| + \|y - F(u_h)\| + \|F(u_i) - F(u_h) - F'(u_i)(u_i - u_h)\|)],$$

hence, applying the tangential cone condition (2.3)

$$S_F \leq 2\nu \sum_{i=h}^{j-1} \|y - F(u_i)\|^2 + (1 + \nu) \sum_{i=h}^{j-1} \|y - F(u_i)\| \|y - F(u_h)\|.$$

From (2.19), we finally find

$$S_F \leq (1 + 3\nu) \sum_{i=h}^{j-1} \|y - F(u_i)\|^2. \quad (2.23)$$

*Term  $S_{\hat{A}}$ :* Utilizing the hypotheses (2.6) and (2.9b), with  $\delta = 0$ , we have

$$\begin{aligned} S_{\hat{A}} &\leq \sum_{i=h}^{j-1} \lambda_i \left| \left\langle y - \hat{A}(u_i), \hat{A}'(u_i)(u_h - u^\dagger) \right\rangle \right| \leq \sum_{i=h}^{j-1} \lambda_i \|y - \hat{A}(u_i)\| \|\hat{A}'(u_i)(u_h - u^\dagger)\| \\ &\leq L_{\hat{A}} C_{\hat{A}}^0 \rho \sum_{i=h}^{j-1} \lambda_i \leq L_{\hat{A}} C_{\hat{A}}^0 C_{\lambda}^0 \rho \sum_{i=h}^{j-1} \|F(u_i) - y\|^2 \end{aligned} \quad (2.24)$$

Inequalities (2.24) and (2.23), inserted in Equation (2.22), give

$$|\langle e_h - e_j, e_h \rangle| \leq (1 + 3\nu + L_{\hat{A}} C_{\hat{A}}^0 C_{\lambda}^0 \rho) \sum_{i=h}^{j-1} \|y - F(u_i)\|^2. \quad (2.25)$$

Reasoning in the same way for  $|\langle e_h - e_k, e_h \rangle|$  we find

$$|\langle e_h - e_k, e_h \rangle| \leq (1 + 3\nu + L_{\hat{A}} C_{\hat{A}}^0 C_{\lambda}^0 \rho) \sum_{i=k}^{h-1} \|y - F(u_i)\|^2. \quad (2.26)$$

From these estimates, it follows that both  $|\langle e_h - e_j, e_h \rangle|$  and  $|\langle e_h - e_k, e_h \rangle|$  go to zero as  $k \rightarrow \infty$  thanks to (2.16). Therefore, from (2.21) and (2.20), we find that  $\{e_k\}$  is a Cauchy sequence. Then, from (2.18), we derive that  $\{u_k\}$  is a Cauchy sequence as well hence the assertion of the theorem follows.  $\square$

In the case of noisy data, we cannot expect that the iteratively regularized Landweber iteration converges, since  $y^\delta$  might not belong to the range of the operator  $F$ . In fact, we can only obtain a stable approximation of a solution of  $F(u) = y$  provided the discrepancy principle (2.2) is employed, i.e, the iteration is stopped after a finite number of steps. Here, assuming that the discrepancy principle (2.2) holds, we prove a stability result for the iteratively regularized Landweber iteration in the case of noisy data.

**Theorem 2.6** *Under Assumption 2.1, (2.8) and (2.9), let  $k_* = k_*(\delta, y^\delta)$  be chosen such that (2.2) holds. Moreover, we assume that for each fixed iteration  $k$ ,  $\lambda_k^\delta \rightarrow \lambda_k$ , as  $\delta \rightarrow 0$ . Then, the data driven iteratively regularized Landweber iteration (2.1) converges to a solution of  $F(u) = y$ , as  $\delta \rightarrow 0$ .*

*Proof:* We recall that  $u^\dagger$  represents the limit of the iteration (2.1) when  $\delta = 0$ , i.e, the case with data  $y$ . Let  $\{\delta_n\}$  be a sequence such that  $\{\delta_n\} \rightarrow 0$  as  $n \rightarrow \infty$  and denote with  $y_n := y^{\delta_n}$  the sequence of the perturbed data. Let  $k_n = k_*(\delta_n, y_n)$  be the stopping index for which the discrepancy principle (2.2) holds, i.e.

$$\left\| F(u_{k_n}^{\delta_n}) - y_n \right\| \leq \tau \delta_n. \quad (2.27)$$

We now distinguish two cases as  $n \rightarrow \infty$ :

- (i)  $k_n \rightarrow k$ ,  $k \in \mathbb{R}^+$ ;
- (ii)  $k_n \rightarrow +\infty$ .

*Case (i):* to avoid technicalities, we can assume that  $k_n = k$ , for all  $n \in \mathbb{N}$ . Therefore, from (2.27) we get

$$\left\| F(u_k^{\delta_n}) - y_n \right\| \leq \tau \delta_n.$$

From the continuity hypotheses on  $F$ ,  $F'$  and  $\hat{A}$ ,  $\hat{A}'$  and recalling (2.9b), we find that

$$u_k^{\delta_n} \rightarrow u_k, \quad F(u_k^{\delta_n}) \rightarrow F(u_k) = y, \quad \hat{A}(u_k^{\delta_n}) \rightarrow \hat{A}(u_k), \quad \lambda_k^{\delta_n} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

This means that the  $k$ -th iterate of (2.1) is a solution of  $F(u) = y$  and thus the iteration terminates with  $u^\dagger = u_k$ .

*Case (ii):* We choose  $k$  sufficiently large such that  $k_n > k$  and by Proposition 2.3 we have

$$\left\| u_{k_n}^{\delta_n} - u^\dagger \right\| \leq \left\| u_k^{\delta_n} - u^\dagger \right\| \leq \left\| u_k^{\delta_n} - u_k \right\| + \left\| u_k - u^\dagger \right\| \quad (2.28)$$

Given  $\epsilon > 0$ , from Theorem 2.5 there exists  $\bar{k} = \bar{k}(\epsilon)$  such that

$$\left\| u_k - u^\dagger \right\| < \epsilon/2, \quad \forall k > \bar{k}. \quad (2.29)$$

Moreover, for  $k$  fixed, there exists  $\bar{n} = \bar{n}(\epsilon, k)$  such that for all  $n > \bar{n}$

$$\left\| u_k^{\delta_n} - u_k \right\| < \epsilon/2, \quad \forall n > \bar{n}. \quad (2.30)$$

In fact, writing explicitly the difference  $u_k^{\delta_n} - u_k$ , we get

$$\begin{aligned} u_k^{\delta_n} - u_k &= u_{k-1}^{\delta_n} - u_{k-1} - F'(u_{k-1}^{\delta_n})^*(F(u_{k-1}^{\delta_n}) - y_n) + F'(u_{k-1})^*(F(u_{k-1}) - y) \\ &\quad - \lambda_{k-1}^{\delta_n} \hat{A}'(u_{k-1}^{\delta_n})^*(\hat{A}(u_{k-1}^{\delta_n}) - y_n) + \lambda_{k-1} \hat{A}'(u_{k-1})^*(\hat{A}(u_{k-1}) - y). \end{aligned}$$

The estimate (2.30) follows using again the continuity property of  $F$  and  $\hat{A}$ , and also the hypothesis that  $\lambda_{k-1}^{\delta_n} \rightarrow \lambda_{k-1}$ , for every fixed iteration. Therefore, by (2.30) and (2.29) in (2.28), we find  $\left\| u_{k_n}^{\delta_n} - u^\dagger \right\| < \epsilon$  for  $n > \bar{n}$ , which means that  $u_{k_n}^{\delta_n} \rightarrow u^\dagger$  as  $n \rightarrow \infty$ .  $\square$

### 3. NUMERICAL EXPERIMENTS

In this section we present some numerical examples related to the iteration (2.1) both for linear operator equations and nonlinear ones. We first consider the linear case for which the operator is the Radon transform. Then, we present some results for the Schlieren model which we take as the prototype for nonlinear operators.

In the following numerical examples, we always consider noise in the data hence we do not need distinguish anymore between  $\lambda_k$  and  $\lambda_k^\delta$ . For this reason, in order to simplify the notation, from this point on we merely use the notation  $\lambda_k$  instead of  $\lambda_k^\delta$ .

All the numerical results in the following sections are based on the assumption that the damping term in (2.1) is a bounded linear map  $\tilde{A} : X \rightarrow Y$ , hence the iterates become

$$u_{k+1} := u_k - F'(u_k)^*(F(u_k) - y^\delta) - \lambda_k \tilde{A}^*(\tilde{A}u_k - y^\delta), \quad k \in \mathbb{N}. \quad (3.1)$$

To build the operator  $\tilde{A} : X \rightarrow Y$ , we assume to have some a-priori information about (1.1) in the form of a finite set of expert data that is

$$(u^{(i)}, y^{(i)}) \in X \times Y, \text{ for } i = 1, \dots, n, \quad (3.2)$$

where  $n > 0$ , and we define, for  $\tilde{T} \in B(X, Y)$ , the functional

$$l(\tilde{T}) = \frac{1}{2} \sum_{i=1}^n \left\| \tilde{T}u^{(i)} - y^{(i)} \right\|_Y^2. \quad (3.3)$$

The bounded operator  $\tilde{A}$  is defined as the operator which minimizes the functional (3.3), i.e.,

$$l(\tilde{A}) = \min_{\tilde{T} \in \mathcal{B}(X, Y)} l(\tilde{T}) = \min_{\tilde{T} \in \mathcal{B}(X, Y)} \left[ \frac{1}{2} \sum_{i=1}^n \left\| \tilde{T}u^{(i)} - y^{(i)} \right\|_Y^2 \right]. \quad (3.4)$$

Let  $H \in B(X, Y)$  and  $t \in J$  where  $J \subset \mathbb{R}$  is an interval containing the origin. Then

$$\begin{aligned} \frac{l(\tilde{T} + tH) - l(\tilde{T})}{t} &= \frac{1}{2t} \left\{ \sum_{i=1}^n \left[ \left\langle (\tilde{T} + tH)u^{(i)} - y^{(i)}, (\tilde{T} + tH)u^{(i)} - y^{(i)} \right\rangle - \left\langle \tilde{T}u^{(i)} - y^{(i)}, \tilde{T}u^{(i)} - y^{(i)} \right\rangle \right] \right\} \\ &= \sum_{i=1}^n \left[ \left\langle \tilde{T}u^{(i)} - y^{(i)}, Hu^{(i)} \right\rangle + \frac{t}{2} \left\langle Hu^{(i)}, Hu^{(i)} \right\rangle \right]. \end{aligned}$$

Therefore, as  $t \rightarrow 0$ , for each  $H \in B(X, Y)$  and every  $(u^{(i)}, y^{(i)}) \in X \times Y$ , for  $i = 1, \dots, n$ , the operator  $\tilde{A}$  is defined as the operator which satisfies the condition

$$\sum_{i=1}^n \left\langle \tilde{A}u^{(i)} - y^{(i)}, Hu^{(i)} \right\rangle = 0, \quad \forall H \in \mathcal{B}(X, Y)$$

which implies that

$$\sum_{i=1}^n \tilde{A}u^{(i)} = \sum_{i=1}^n y^{(i)}. \quad (3.5)$$

Now, denoting by  $\{e_k(x)\}_{k \in \mathbb{N}}$  and  $\{\bar{e}_j(y)\}_{j \in \mathbb{N}}$  complete orthonormal families of  $X$  and  $Y$ , respectively, we have

$$\begin{aligned} u^{(i)} &= \sum_{k \in \mathbb{N}} \langle u^{(i)}, e_k \rangle e_k = \sum_{k \in \mathbb{N}} u_k^{(i)} e_k, \quad \forall i = 1, \dots, n, \\ y^{(i)} &= \sum_{j \in \mathbb{N}} \langle y^{(i)}, \bar{e}_j \rangle \bar{e}_j = \sum_{j \in \mathbb{N}} y_j^{(i)} \bar{e}_j, \quad \forall i = 1, \dots, n, \end{aligned} \quad (3.6)$$

and we recall that each bounded operator between Hilbert spaces has a matrix representation given by

$$\tilde{A}e_k = \sum_{j \in \mathbb{N}} \tilde{a}_{jk} \bar{e}_j, \quad \text{where } \tilde{a}_{jk} = \langle \tilde{A}e_k, \bar{e}_j \rangle, \quad (3.7)$$

see for example [10]. In this way, from (3.6) and (3.7), we find

$$\tilde{A}u^{(i)} = \sum_{k \in \mathbb{N}} u_k^{(i)} \tilde{A}e_k = \sum_{k, j \in \mathbb{N}} \tilde{a}_{jk} u_k^{(i)} \bar{e}_j, \quad \forall i = 1, \dots, n.$$



Therefore, Equation 3.5 can be rewritten as

$$\sum_{i=1}^n \sum_{k,j \in \mathbb{N}} \tilde{a}_{jk} u_k^{(i)} \bar{e}_j = \sum_{i=1}^n \sum_{j \in \mathbb{N}} y_j^{(i)} \bar{e}_j$$

that is, for each  $j \in \mathbb{N}$  and  $i = 1, \dots, n$ , we have

$$\sum_{k \in \mathbb{N}} \tilde{a}_{jk} u_k^{(i)} = y_j^{(i)}.$$

In the sequel we approximate the matrix  $\tilde{a}_{jk}$ , with a matrix  $A \in \mathbb{R}^{M \times N}$ , that is we take a finite number of elements of the orthonormal families of the Hilbert spaces  $X$  and  $Y$ . To be more precise, we employ  $\{e_k\}$ , for  $k = 1, \dots, N$ , and  $\{\bar{e}_j\}$ , for  $j = 1, \dots, M$ .

**Construction of A.** Following considerations above, we consider

$$(u^{(i)}, y^{(i)}) \in \mathbb{R}^N \times \mathbb{R}^M, \text{ for } i = 1, \dots, n, \quad (3.8)$$

where  $n > 0$ . With this choice, the matrix  $A \in \mathbb{R}^{M \times N}$  is defined as the linear map such that

$$AU = Y, \quad (3.9)$$

where  $U \in \mathbb{R}^{N \times n}$  and  $Y \in \mathbb{R}^{M \times n}$  are the matrices which contain columnwise the data  $u^{(i)}$  and  $y^{(i)}$ , respectively, for  $i = 1, \dots, n$ . The matrix  $A$  can be obtained by utilizing a Singular Value Decomposition (SVD) on  $U$ , hence

$$A = YU^\dagger. \quad (3.10)$$

**Remark 3.1** *One of the main issue of this approach is the computational burden in building the matrix  $A$  that is the amount of the storage needed for the data. In fact the dimensions  $M$  and  $N$  of the matrix  $A$  could be very large even for small scale inverse problems.*

The following numerical experiments have been accomplished using Matlab and some of its functions.

**3.1. Linear Operator - Radon Transform.** In this section, we present and discuss some numerical experiments assuming that the operator  $F$  in (1.1) is the Radon transform, i.e., we consider the operator equation

$$Ru = y. \quad (3.11)$$

We refer, for instance, the reader to [12, 14] for the definition and properties of the Radon transform.

In the sequel, we implement the Data-Driven Iteratively Regularized Landweber Iteration (DDIRLI) (3.1), adapted to Equation (3.11), i.e

$$u_{k+1} := u_k - \omega_R R^* (Ru_k - y^\delta) - \lambda_k^\delta \tilde{A}^* (\tilde{A}u_k - y^\delta), \quad k \in \mathbb{N}, \quad (3.12)$$

where  $R^*$  is the backprojection operator and  $\omega_R$  is a positive constant such that  $\|R^*(Ru_k - y^\delta)\| \leq 1$ . The operator  $\tilde{A}$  can be realized by a matrix  $A$  which is generated following the approach discussed at the beginning of Section 3. For this purpose, we use the following set of training data and parameters:

- (a) 180 different equally distributed angles  $\theta$  within the interval  $[0, \pi)$ ; specifically, we consider the set  $\Theta = \left\{0, \frac{\pi}{180}, \frac{\pi}{90}, \dots, \frac{179\pi}{180}\right\}$ ;
- (b) A sample of 50 grayscale images of handwritten digits, from MNIST database, and their related sinograms, which are obtained using the 180 directions of (a).

Matrices  $U$  and  $Y$  in (3.9) are built utilizing data  $(u^{(i)}, y^{(i)})$ , where  $u^{(i)}$  and  $y^{(i)}$  are, respectively, the images and their related sinograms described in (b). The number of training pairs utilized for the creation of  $A$  will be specified in each numerical test.

In the implementation, for the set of parameters we make these choices:

(i) We choose

$$\omega_R = 10^{-2} \quad \text{and} \quad \lambda_k = 7.7 \times 10^{-5} \|Ru_k - y^\delta\|_2^2, \quad (3.13)$$

where  $\|\cdot\|_2$  is the spectral norm. The value of  $\lambda_k$  is inferred by conditions (2.9).

- (ii) As stopping rule we use the discrepancy principle (2.2), where the choice of  $\tau$  will be specified for each test (see tables below), or a maximum number of 100 iterates;
- (iii) The synthetic data  $y^\delta$  are generated by adding a gaussian distributed noise with zero mean and variance  $\sigma^2$ , specified at each test, to the matrix of the exact data  $y$ ;
- (iv) The initial guess  $u_0$  is always set to be  $u_0 = 0$ .

Results of (3.12) are compared with the outcomes of the Landweber iteration, that is

$$u_{k+1} := u_k - \omega_R R^*(Ru_k - y^\delta), \quad k \in \mathbb{N}, \quad (3.14)$$

and of the Iteratively Regularized Landweber Iteration (IRLI) given by

$$u_{k+1} := u_k - \omega_R R^*(Ru_k - y^\delta) - \beta_k(u_k - u^{(0)}), \quad k \in \mathbb{N}, \quad (3.15)$$

where, according to the parameters rules in [9, Section 3.2], we choose  $\beta_k = (\frac{1}{4})^{k+1}$ , for  $k \in \mathbb{N}$ , and we use as  $u^{(0)}$  a noise version of the image to be reconstructed, see Figure 1. Finally, the numerical results provided by these iterative algorithms are compared with those of the Filtered Back-Projection (FBP) algorithm, which performs the inverse Radon transform. For this last case, we run the Matlab function “iradon” making the choice of the Ram-Lak filter.

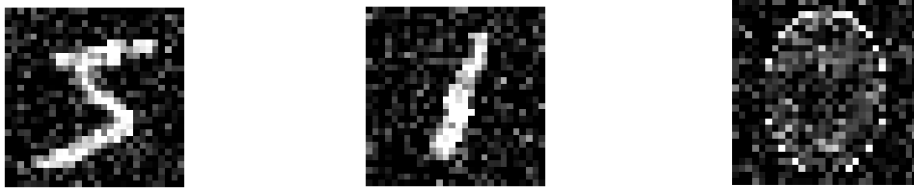


FIGURE 1. Choices of  $u^{(0)}$  in (3.15): image on the left is used in Test 1 and Test 3a. Image in the center is used in Test 2a/2b and Test 3b. Image on the right is used in Test 4.

We used the training and validation sets of the MNIST dataset of handwritten digits, see [13]. We essentially consider four different numerical experiments: In **Test 1** we reconstruct an image which has been utilized to create the matrix  $A$ , i.e., it is contained in the training set of MNIST. In **Test 2a/2b**, using the same matrix  $A$  of test 1, we reconstruct a digit in the validation test of MNIST. In **Test 3a** and **Test 3b**, we reconstruct the same images of Test 1 and Test 2, respectively, but using only partial data, i.e., we consider a cropped noisy version of the sinogram of the true image. We stress that the true image of Test 3b has not been utilized to create the matrix  $A$ . Finally, in **Test 4**, we reconstruct the Shepp Logan phantom in order to test our method on an image which is completely different from the ones in the training and validation sets.

The choices and values of the main parameters, which are utilized for the generation of the data and to run the schemes (3.12), (3.14) and (3.15), are contained in the tables below, i.e., Tables 1, 2, 3, 4, 5 and 6: In the left part of each table, we specify the choice of the variance in the gaussian noise, the value of  $\delta$  and  $\tau$ . In the right part we state some of the results of the numerical test, i.e., the total number of iterates before one of the two stopping criteria in (ii) has been accomplished, the computational time and, finally, the relative error  $\|u_{\text{true}} - u_{\text{rec}}\|_2 / \|u_{\text{true}}\|_2$ , where  $u_{\text{true}}$  is the image to be reconstructed while  $u_{\text{rec}}$  is the reconstructed image. We stress that the execution time of DDIRLI is the result of the construction of the matrix  $A$  and the iterates of (3.12).

**Test 1.** (Figure 2). *Target: reconstruct one of the image from the training set which has been used to create the matrix  $A$ .* See Figure 2 for the “True Image”. The initial datum  $y^\delta$  is given by the sinogram of the true image  $y$ , adding a gaussian distributed noise of zero mean and variance  $\sigma^2 = 0.5$ . In the discrepancy principle (2.2) we choose  $\tau = 1.1$ . See Table 1. We observe, from Figure 2 and Figure 3, how the presence of  $A$  (and the fact that the information of the “True Image” are included in  $A$ ) “helps” the iterates (3.12) in recognizing the true image, as expected. Indeed, despite the presence of a strong noise, DDIRLI provides better results compared to those of the Landweber iteration (3.14) and IRLI (3.15), with a smaller number of iterates. Note that the initial guess  $u^{(0)}$  in IRLI, see (3.15), is the “True Image” corrupted with a gaussian noise of variance 0.05 only.

**Test 2a/2b.** (Figures 4 and 6). *Target: reconstruct one of the image from the validation set, not utilized*

to create  $A$ . We reconstruct a digit from the validation set of MNIST. The difference between the two tests, see Figures 4 and 6, is the number of the samples from training data used to create the matrix  $A$ . In fact, we used in Test 2a only 50 images while in Test 2b 150 images. From these numerical examples, we clearly observe an improvement of the results of DDIRLI when we increase the number of images utilized to create the matrix  $A$ . In fact, Test 2b shows very good results, comparable with those of the Landweber scheme and IRLI.

**Test 3a/3b** (Figures 8 and 10). *Target: reconstruct images from the training or the validation set using partial data.* We consider the case of partial noisy measurements which corresponds to taking a cut version of the sinogram  $y^\delta$  with a gaussian distributed noise of zero mean and variance  $\sigma = 0.03$ . The cropped sinogram is obtained maintaining information only in some directions and setting to zero all the rest. See Figure 8 and Figure 10.

In Test 3a, we reconstruct an image belonging to the training set and which has been utilized to create the matrix  $A$ . On the contrary, in Test 3b, we reconstruct an image from the validation set and therefore was not utilized to generate the matrix  $A$ . The outcomes of our method are promising in both the cases.

**Test 4** (Figure 12 and 13). *Target: reconstruct Shepp Logan phantom, i.e., an image which is completely different from the elements contained in training data which are used to create the matrix  $A$ .*

To use the MNIST training set to create  $A$ , the Shepp Logan image has been down-sampled to  $28 \times 28$ . In numerical simulation, we observe that, in order to get a good reconstruction of the true image and comparable with the numerical results of the other methods, we use 600 images from the training set.

We stress that Figures 3 C), 5 C), 7 C), 9 C) and 11 C) show the ratios between the residuals of the data driven model,  $\|Au_k - y^\delta\|$ , and the model driven approach,  $\|Ru_k - y^\delta\|$ . A value greater than 1 of the ratio implies that the data driven approach has a significant influence on the iteration process. In fact, since the initial guess is  $u_0 = 0$ , from plots C), we observe how our scheme, in the first steps, is giving more weight to the information contained in  $A$  to reconstruct the object. Figures 3 D), 5 D), 7 D), 9 D) and 11 D) enlight how fast the product  $\|Ru_k - y^\delta\|_2 \|Au_k - y^\delta\|_2$  is going to zero. In fact, since  $\lambda_k = 7.7 \times 10^{-5} \|Ru_k - y^\delta\|_2^2$ , we have that

$$\frac{\|Ru_k - y^\delta\|_2}{\lambda_k \|Au_k - y^\delta\|_2} = \frac{1}{7.7 \times 10^{-5} \|Ru_k - y^\delta\|_2 \|Au_k - y^\delta\|_2}.$$

Finally, we stress that plots A), B) and D) in Figures 3, 5, 7, 9, 11 are all in logarithmic scale.

TABLE 1. Test 1. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	0.5	13,36110296	1.1	25	1,0732974	0,107250035
IRLI				58	1,3639758	0,114869088
LANDWEBER				73	1,734035	0,12467763
FBP		/	/	/	0,0546157	0,124274036

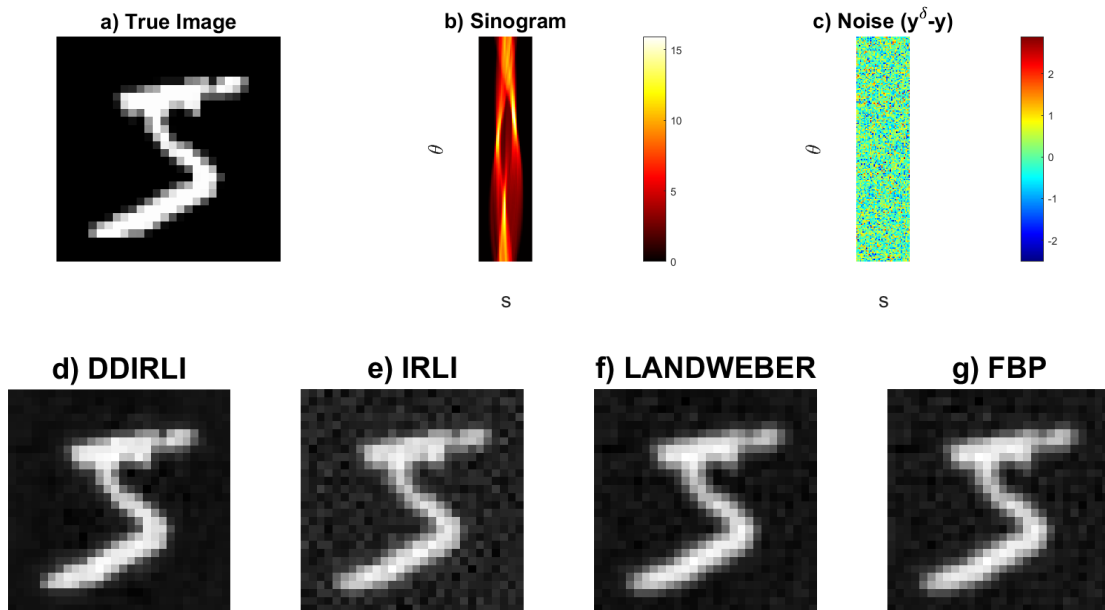


FIGURE 2. Test 1. Reconstruction of an image from the training set which has been used to create  $A$  in (3.10). a) Image to be reconstructed; b) Sinogram  $y$  of the true image; c) Plot of the gaussian noise, i.e.,  $y^\delta - y$ ; d)-g) Reconstructions by the different methods. DDIRLI gives a very good reconstruction, with more details compared to IRLI, Landweber and FBP.

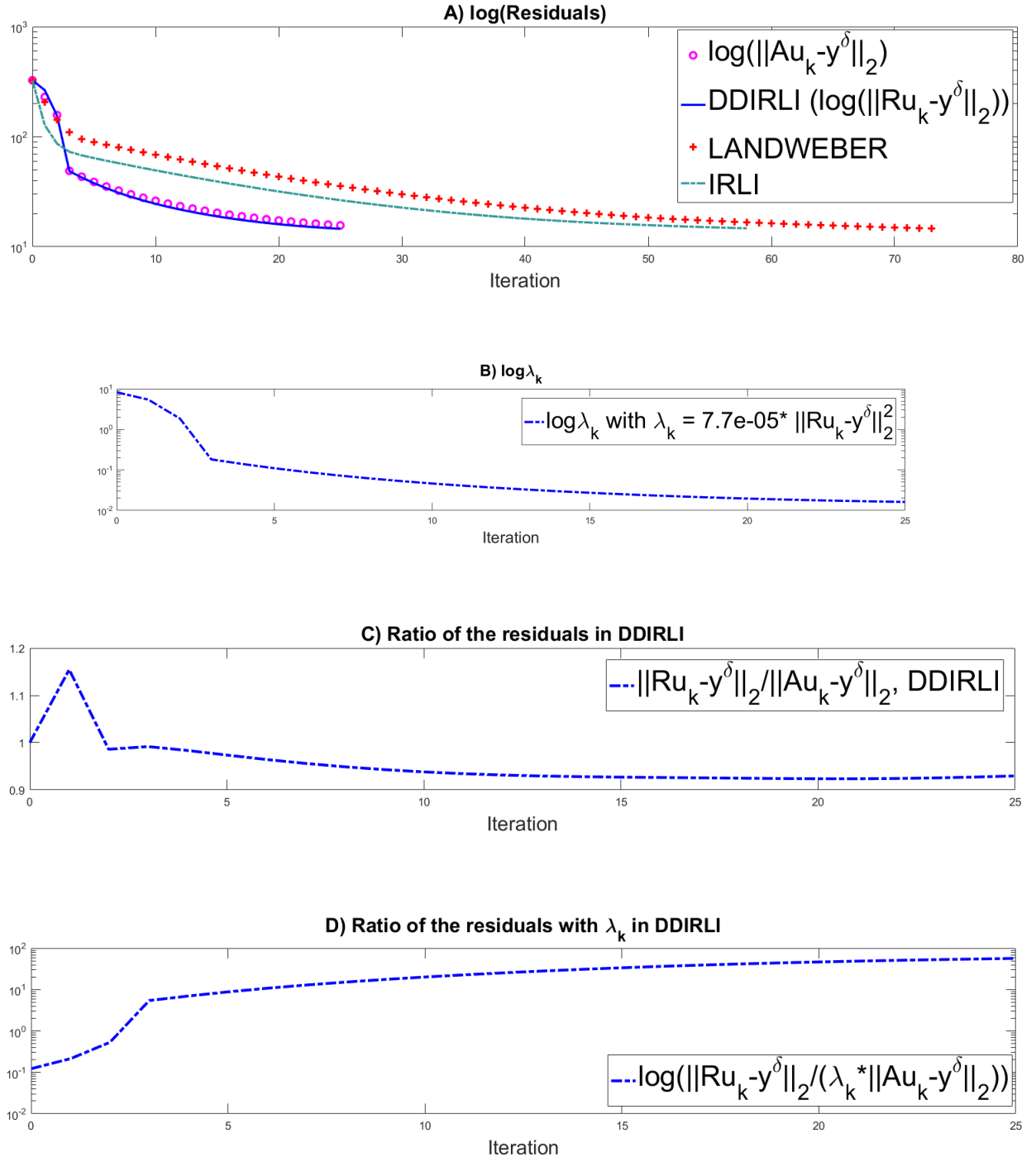


FIGURE 3. Test 1. A) Residual in logarithmic scale at each step implementing (3.12), (3.14) and (3.15); B) plot of  $\log(\lambda_k)$ ; C) plot of the ratio  $\|Ru_k - y^\delta\|_2 / \|Au_k - y^\delta\|_2$ ; D) plot in logarithmic scale of the ratio  $\|Ru_k - y^\delta\|_2 / (\lambda_k \|Au_k - y^\delta\|_2)$ .

TABLE 2. Test 2a. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	0.5	13,41600448	1.1	39	1,474326	0,203051179
IRLI				52	1,2946457	0,139542084
LANDWEBER				67	1,743846	0,14349784
FBP		/	/	/	0,0273525	0,138023154

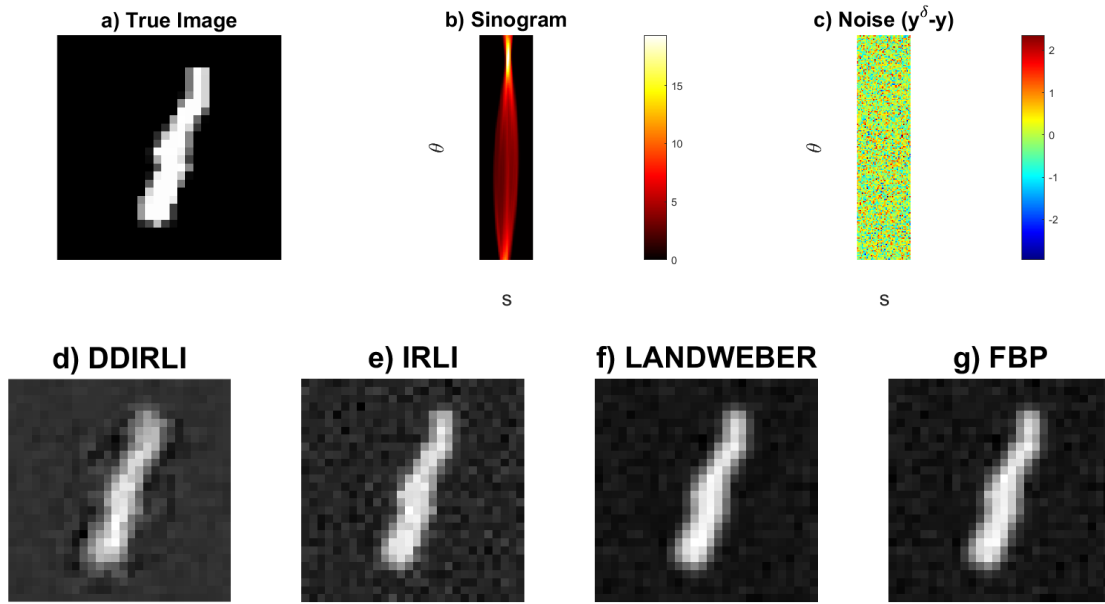


FIGURE 4. Test 2a. Reconstruction of an image from the validation set not used to create  $A$  in (3.10). Only a sample of 50 images from the training set has been used to build  $A$ . a) Image to be reconstructed; b) Sinogram ( $y$ ) of the true image; c) Plot of the gaussian noise, i.e.,  $y^\delta - y$ ; d)-g) Reconstructions by the different methods. IRLI, Landweber and FBP give better results than the DDIRLI method. We stress that the discrepancy principle for DDIRLI is satisfied after only 39 iterations.

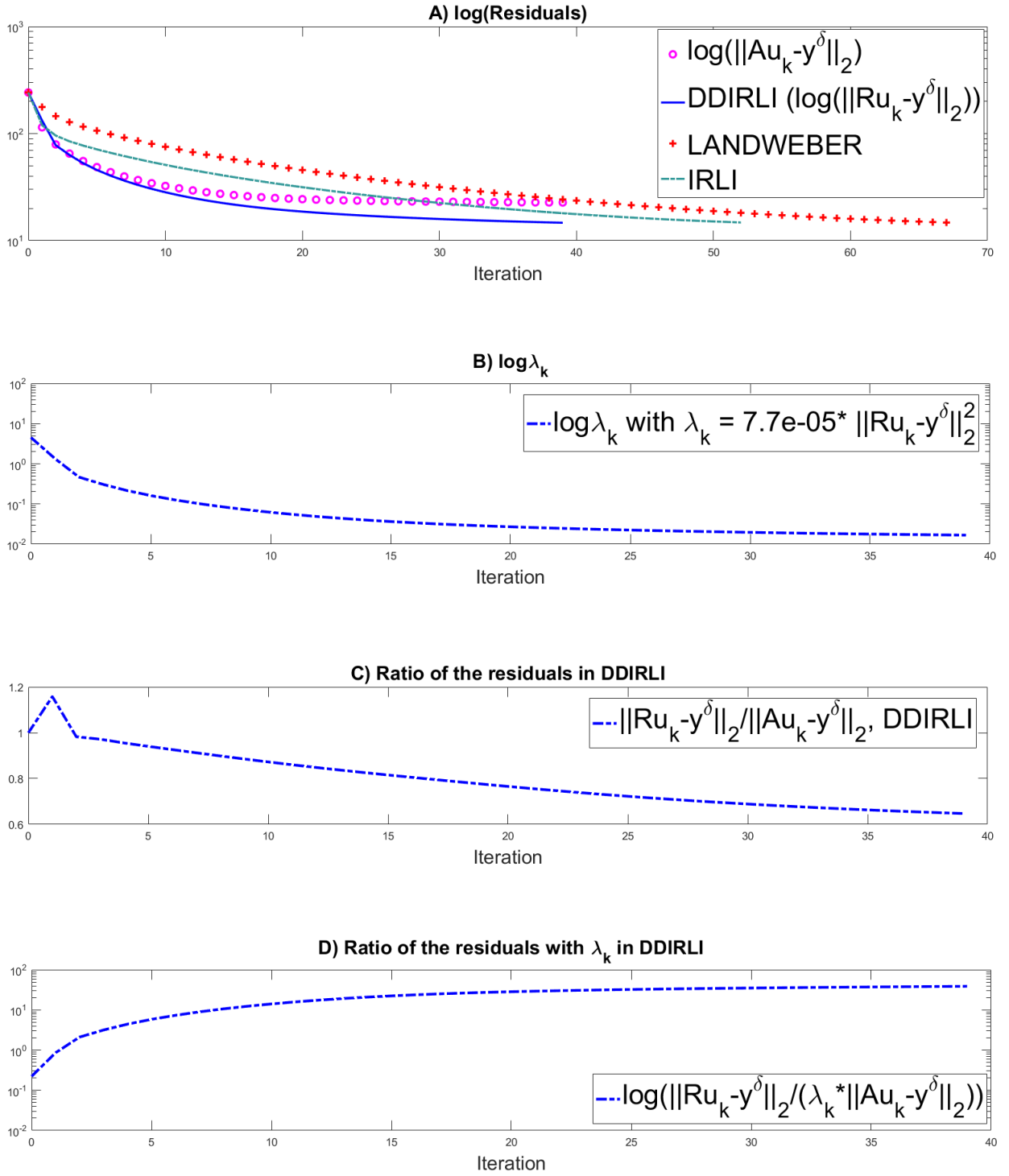


FIGURE 5. Test 2a. A) Residual in logarithmic scale at each step implementing (3.12), (3.14) and (3.15); B) plot of  $\log(\lambda_k)$ ; C) plot of the ratio  $\|Ru_k - y^\delta\|_2 / \|Au_k - y^\delta\|_2$ ; D) plot in logarithmic scale of the ratio  $\|Ru_k - y^\delta\|_2 / (\lambda_k \|Au_k - y^\delta\|_2)$ .

TABLE 3. Test 2b. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	0.5	13,71978428	1.1	24	1,0422007	0,161019463
IRLI				53	1,3471281	0,135825795
LANDWEBER				66	1,5579816	0,142645348
FBP		/	/	/	0,0419196	0,135176876

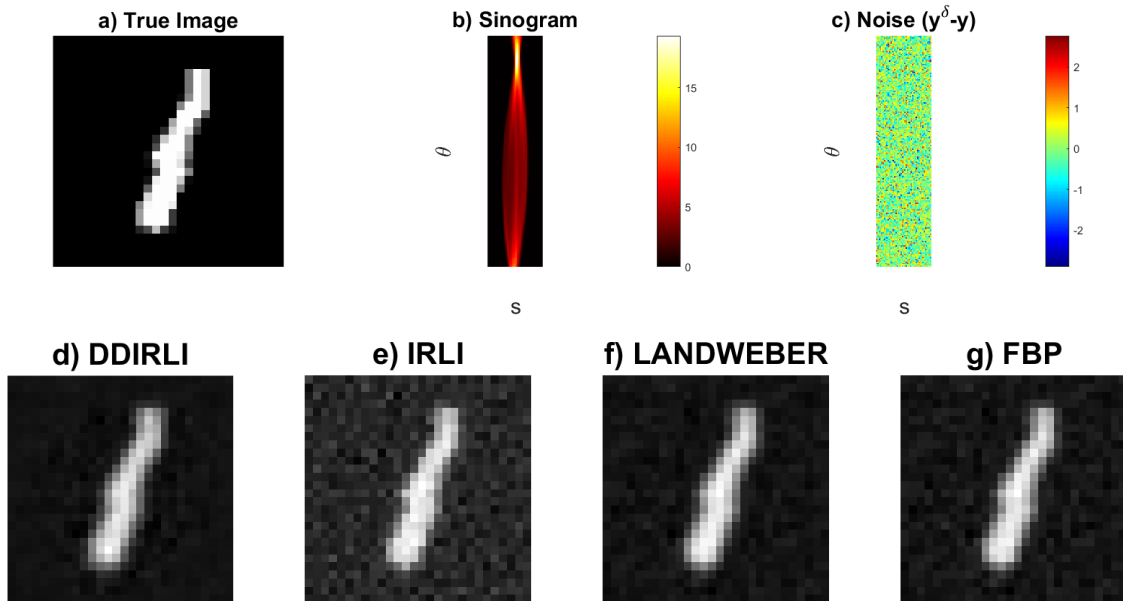


FIGURE 6. Test 2b. Reconstruction of an image from the validation set not used to create  $A$  in (3.10). Only a sample of 150 images from the training set has been used to build  $A$ . a) Image to be reconstructed; b) Sinogram ( $y$ ) of the true image; c) Plot of the gaussian noise, i.e.,  $y^\delta - y$ ; d)-g) Reconstructions by the different methods. Reconstructions are all comparable. We stress that the discrepancy principle for DDIRLI is satisfied after only 24 iterations.



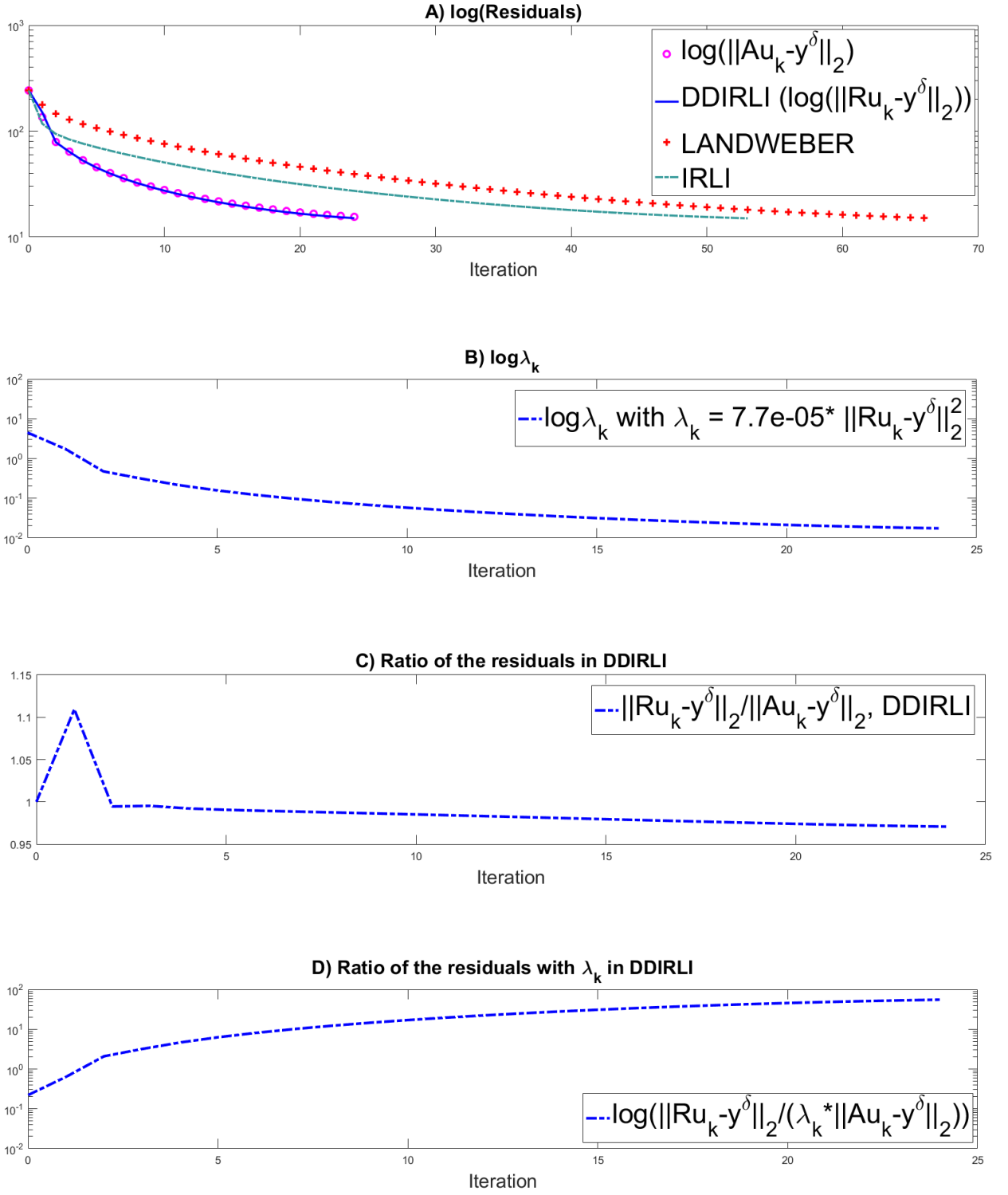


FIGURE 7. Test 2b. A) Residual in logarithmic scale at each step implementing (3.12), (3.14) and (3.15); B) plot of  $\log(\lambda_k)$ ; C) plot of the ratio  $\|Ru_k - y^\delta\|_2 / \|Au_k - y^\delta\|_2$ ; D) plot in logarithmic scale of the ratio  $\|Ru_k - y^\delta\|_2 / (\lambda_k \|Au_k - y^\delta\|_2)$ .

TABLE 4. Test 3a. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	0.03	2,863089254	5	28	1,0690521	0,259157333
IRLI				56	1,1251772	0,236613193
LANDWEBER				90	1,8927882	0,312031414
FBP		/	/	/	0,024378	0,475620179

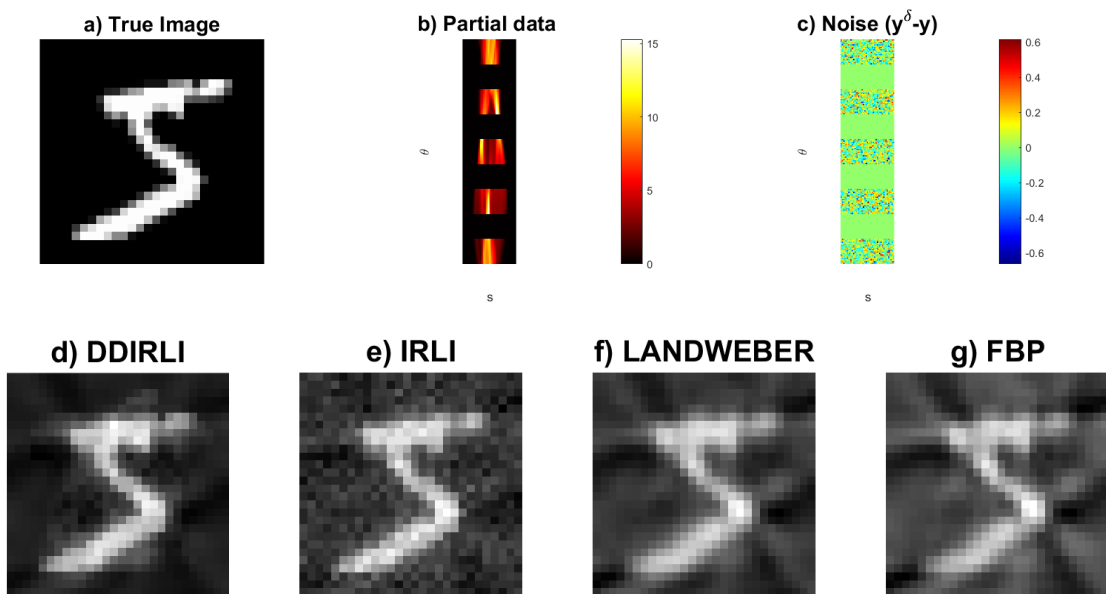


FIGURE 8. Test 3a. Reconstruction of an image from the training set, which has been used to create  $A$ , from partial data. a) Image to be reconstructed; b) Plot of the cropped sinogram of the true image, i.e.,  $y$ ; c) Plot of the gaussian noise, i.e.,  $y^\delta - y$ ; d)-g) Reconstructions by the different methods. DDIRLI and IRLI provide better reconstructions than Landweber and FBP.

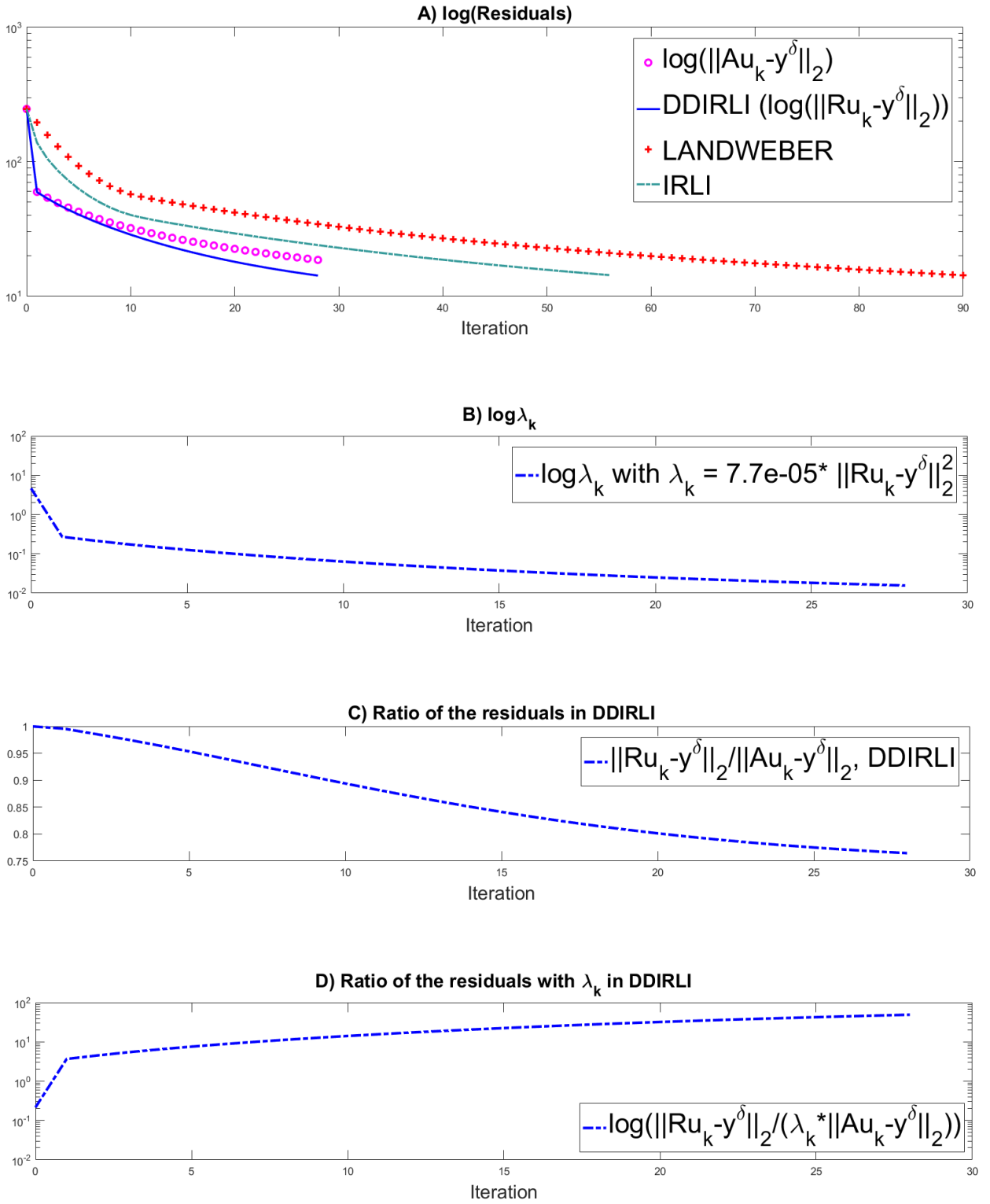


FIGURE 9. Test 3a. A) Residual in logarithmic scale at each step implementing (3.12), (3.14) and (3.15); B) plot of  $\log(\lambda_k)$ ; C) plot of the ratio  $\|Ru_k - y^\delta\|_2 / \|Au_k - y^\delta\|_2$ ; D) plot in logarithmic scale of the ratio  $\|Ru_k - y^\delta\|_2 / (\lambda_k \|Au_k - y^\delta\|_2)$ .

TABLE 5. Test 3b. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	0.03	2,760643462	5	29	1,1999311	0,270013425
IRLI				55	1,1808313	0,240562351
LANDWEBER				80	1,654519	0,289219401
FBP		/	/	/	0,0210105	0,427565297

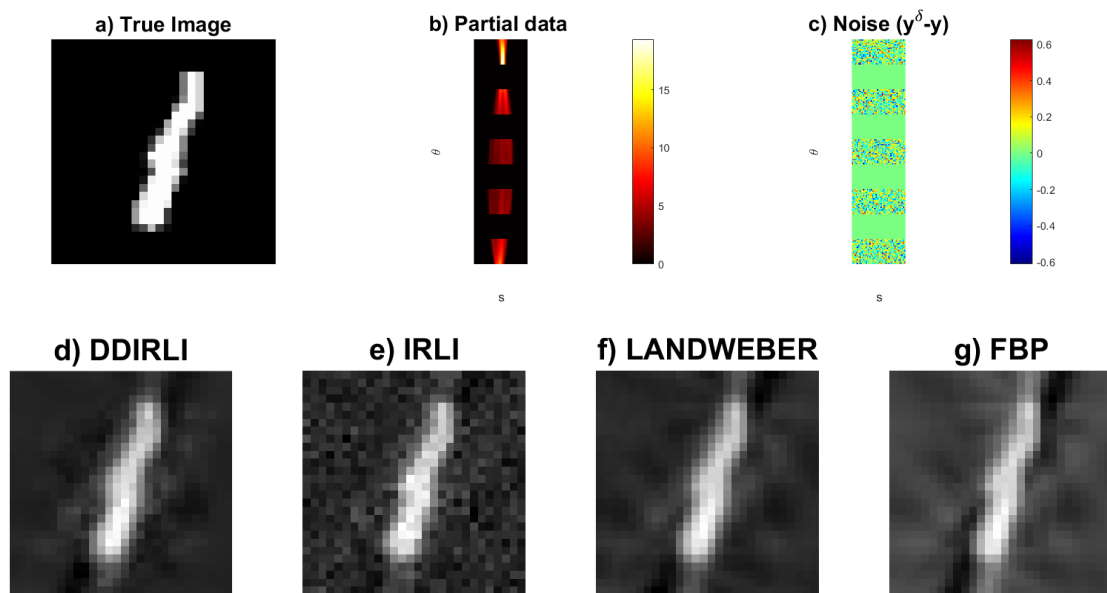


FIGURE 10. Test 3b. Reconstruction of an image from the validation set not used to create  $A$  in (3.10), from partial data. We used the same 150 images of Test 2b to create  $A$ . a) Image to be reconstructed; b) Plot of the cropped sinogram of the true image, i.e.,  $y$ ; c) Plot of the gaussian noise, i.e.,  $y^\delta - y$ ; d)-g) Reconstructions by the different methods.

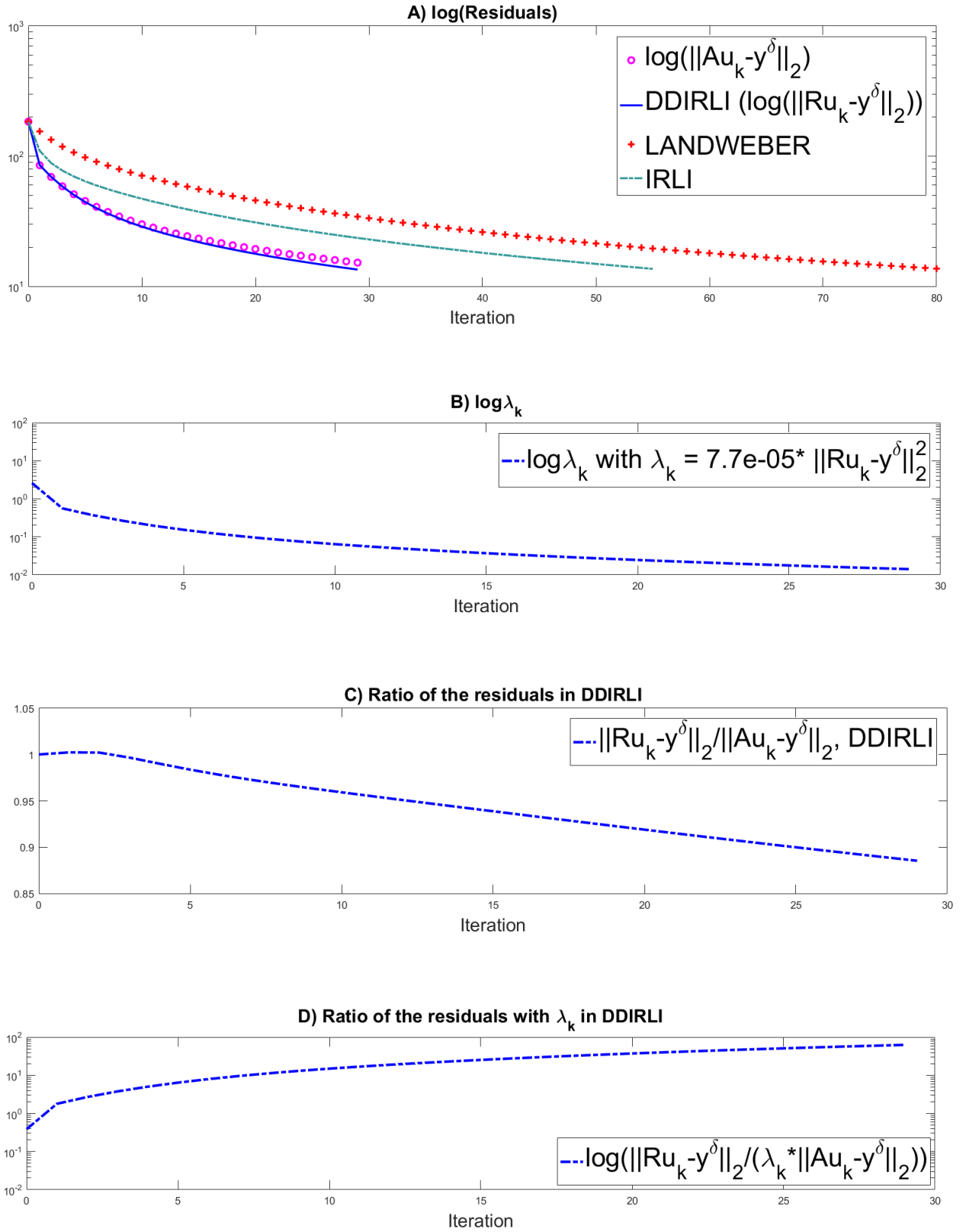


FIGURE 11. Test 3b. A) Residual in logarithmic scale at each step implementing (3.12), (3.14) and (3.15); B) plot of  $\log(\lambda_k)$ ; C) plot of the ratio  $\|Ru_k - y^\delta\|_2 / \|Au_k - y^\delta\|_2$ ; D) plot in logarithmic scale of the ratio  $\|Ru_k - y^\delta\|_2 / (\lambda_k \|Au_k - y^\delta\|_2)$ .

TABLE 6. Test 4. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	0.5	13,56737965	1.1	26	1,4472282	0,399702146
IRLI				34	0,6318276	0,323027126
LANDWEBER				56	1,0820124	0,410787776
FBP		/	/	/	0,0206258	0,382431698

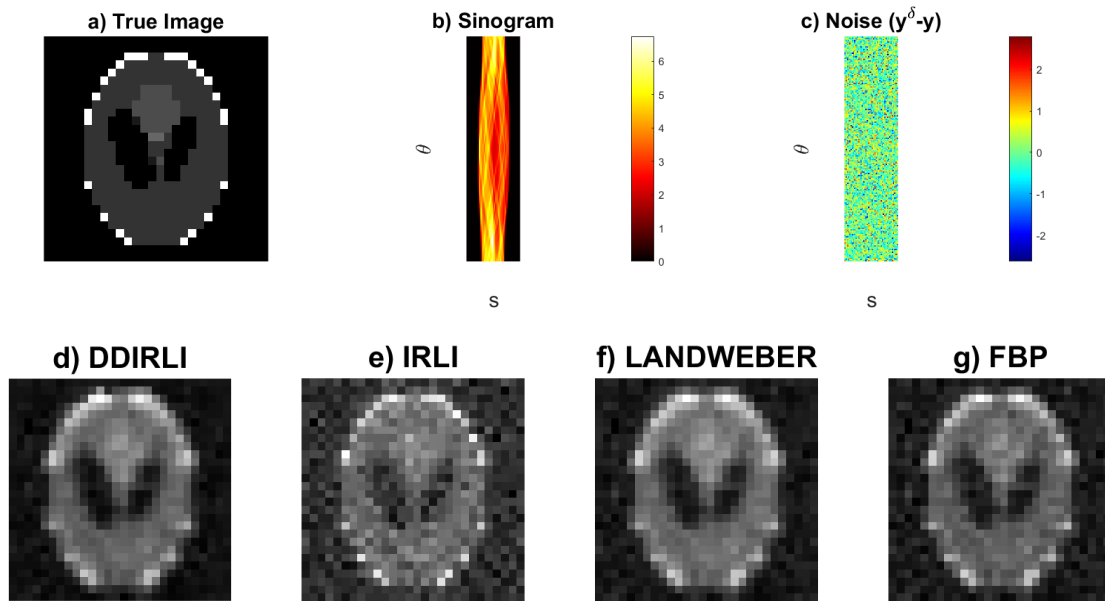


FIGURE 12. Test 4. Reconstruction of the Shepp Logan phantom. An amount of 600 images from the training set have been used to create  $A$ . a) Image to be reconstructed; b) Plot of the cropped sinogram of the true image, i.e.,  $y$ ; c) Plot of the gaussian noise, i.e.,  $y^\delta - y$ ; d)-g) Reconstructions by the different methods.

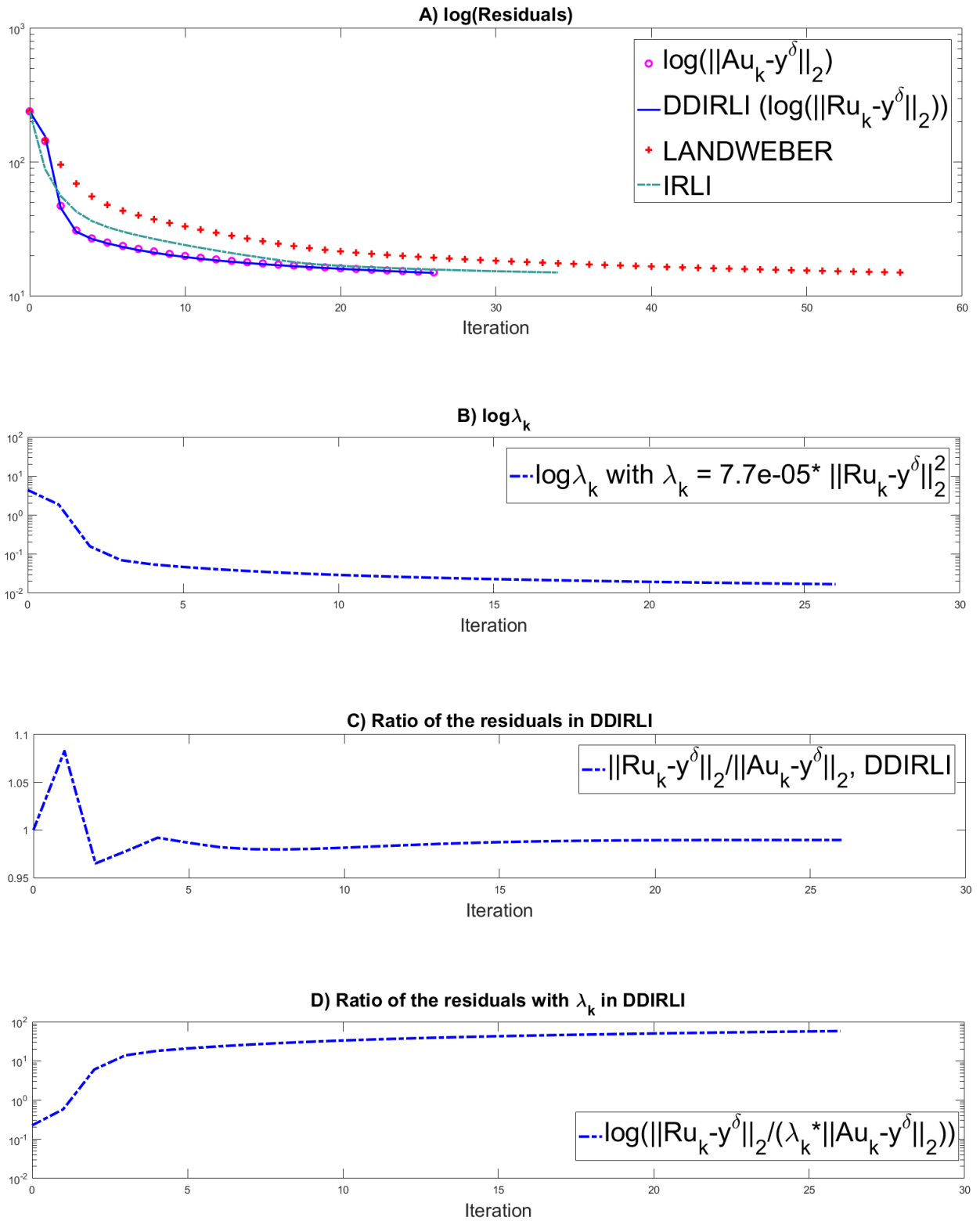


FIGURE 13. Test 4. A) Residual in logarithmic scale at each step implementing (3.12), (3.14) and (3.15); B) plot of  $\log(\lambda_k)$ ; C) plot of the ratio  $\|Ru_k - y^\delta\|_2 / \|Au_k - y^\delta\|_2$ ; D) plot in logarithmic scale of the ratio  $\|Ru_k - y^\delta\|_2 / (\lambda_k \|Au_k - y^\delta\|_2)$ .

**3.2. Nonlinear Operator - Schlieren Model.** We briefly recall the mathematical model behind the Schlieren tomography, which we take as prototype of a nonlinear problem. For the physical model, i.e., the Schlieren optical system and details on data acquisition see, e.g., [5, 11, 16].

Let  $\eta_i \in S^1$ , for  $i = 0, \dots, l-1$  be a set of recording directions, that is

$$\eta_i = \eta(\theta_i) = (\cos \theta_i, \sin \theta_i),$$

where  $\theta_i \in [0, \pi)$ , for  $i = 0, \dots, l-1$ , is the angle. Taking  $B_1(0) \subset \mathbb{R}^2$ , we define the Schlieren operator in the direction of  $\eta_i$  as the square of the Radon transform, i.e.,

$$F_i(u) := R_i^2(u), \quad i = 0, \dots, l-1,$$

where  $R_i$  is the Radon transform along the direction  $\eta_i$ . Physically speaking, the function  $u$  denotes a pressure. To reconstruct  $u$ , we need to solve the system of equations

$$F(u) := (F_0(u), \dots, F_{l-1}(u)) = y^\delta := (y_0^\delta, \dots, y_{l-1}^\delta). \quad (3.16)$$

It was shown in [8] that each operator  $F_i$ , for  $i = 0, \dots, l-1$  is continuous and Fréchet differentiable from  $H_0^1(B_1(0))$  to  $L^2([-1, 1])$  and

$$F_i'(u)h = 2R_i(u)R_i(h), \quad \forall h \in H_0^1(B_1(0)).$$

The adjoint of  $F_i'(u)$  is given by

$$\begin{aligned} F_i'(u)^* : L^2([-1, 1]) &\rightarrow H_0^1(B_1(0)) \\ g &\rightarrow f(g) \end{aligned}$$

with  $f(g)$  the solution to

$$(I - \Delta)f(g) = 2R_i^*(R_i(u)g), \quad (3.17)$$

where  $I$  is the identity operator,  $\Delta$  is the Laplace operator on  $H_0^1(B_1(0))$  and  $R_i^*$  denotes the adjoint of  $R_i$  as operator from  $L^2(B_1(0))$  to  $L^2([-1, 1])$ , i.e.,

$$\begin{aligned} R_i^* : L^2([-1, 1]) &\rightarrow L^2([B_1(0)]) \\ w(x) &\rightarrow w(\langle \eta_i, x \rangle). \end{aligned}$$

We stress that all the numerical simulations of this section are performed without the smoothing operator  $(I - \Delta)$ , contained in (3.17), since both the iteratively regularized Landweber schemes and the Landweber iteration are regularizing.

Therefore, the data driven iteratively regularized Landweber iteration (DDIRLI) for the Schlieren operator is given by

$$u_{k+1} := u_k - \alpha \sum_{i=0}^{l-1} [R_i^*(R_i(u_k)(R_i^2(u_k) - y_i^\delta))] - \lambda_k \tilde{A}^*(\tilde{A}u_k - y^\delta), \quad k \in \mathbb{N}, \quad (3.18)$$

where  $\alpha$  is a relaxation parameter such that the norm of the operator is less or equal to 1. Now, we can create the matrix  $A$  for the Equation (3.16) using the procedure described in Section 3, see equation (3.9). For this purpose, we use the following set of data and parameters:

- (a) 180 different equally distributed angles  $\theta$  within the interval  $[0, \pi)$ ; specifically, we consider the set  $\Theta = \left\{0, \frac{\pi}{180}, \frac{\pi}{90}, \dots, \frac{179\pi}{180}\right\}$ ;
- (b) 10 input images of dimensions  $110 \times 110$  of synthetic pressure functions  $u^{(j)}$ ,  $j = 1, \dots, 10$ , and the related output data. See Figure 14 for the dataset used in numerical Test 1S, 2aS and 2bS. See Figure 15 for the dataset used in numerical Test 3S. Note that the difference between the two datasets is only on the values of the pressure in some pictures.

We create the matrix  $U$  and  $Y$ , as defined in (3.9), utilizing the data in (b).



Again, we compare results of (3.18) with the Landweber scheme, i.e.,

$$u_{k+1} := u_k - \alpha \sum_{i=0}^{l-1} [R_i^*(R_i(u_k)(R_i^2(u_k) - y_i^\delta))], \quad k \in \mathbb{N} \quad (3.19)$$

and with the iteratively regularized Landweber iteration (IRLI)

$$u_{k+1} := u_k - \alpha \sum_{i=0}^{l-1} [R_i^*(R_i(u_k)(R_i^2(u_k) - y_i^\delta))] - \beta_k(u_k - u^{(0)}), \quad k \in \mathbb{N}, \quad (3.20)$$

where, according to the parameters rules in [9, Section 3.2], we choose  $\beta_k = (\frac{1}{4})^{k+1}$ . Here, we specify the hypotheses and the set of parameters which are employed in our numerical experiments for the schemes (3.18), (3.19) and (3.20):

- (i) We choose  $\alpha = 10^{-8}$  and  $\lambda_k = 10^{-7} \|F(u_k) - y^\delta\|_\infty^2$ ;
- (ii) As stopping rule we utilize the discrepancy principle (2.2), where the choice of  $\tau$  will be specified for each test (see tables below), or a maximum number of 400 iterates;
- (iii) Each vector  $y_i^\delta$ , for  $i = 0, \dots, 179$ , of synthetic data is generated by adding a gaussian distributed noise with zero mean and variance  $\sigma^2$ , equal to 100 or 10, to each vector  $y_i$ , for  $i = 0, \dots, 179$ , which represents the exact data;
- (iv) we define  $\delta_S := \max(\|y_i^\delta - y_i\|_2)$ , for  $i = 0, \dots, 179$ ;
- (v) The initial guess is the constant function  $u_0 = 0.01$ ;

We consider three different numerical experiments: in **Test 1S** we reconstruct an image which has been utilized to create the matrix  $A$ . In **Test 2aS/2bS** we propose the reconstruction of an image containing three regions with negative values. In details, in Test 2aS we show the reconstruction adding noise in the data. In Test 2bS we consider the case of partial noisy data. In **Test 3S**, we consider the case where the pressure to be reconstructed is positive in some regions and negative in others.

We stress that the Schlieren imaging has an inherent nonuniqueness issue, since the sign of the true solution is undetermined. This will be clear in the numerical outcomes of the Landweber iteration.

As in the previous section, for each test we provide a table containing, on the left the values of the main parameters utilized in the generation of the data and, on the right some results such as the number of iterations, the computational time and finally the relative error. See Tables 7, 8, 9 and 10.

**Test 1S.** (Figure 16). *Target: reconstruct one of the image which has been utilized to create the matrix  $A$ .* See Figure 16 for the ‘‘True Image’’. All the synthetic data  $y_i^\delta$ , for  $i = 0, \dots, 179$ , are obtained by the real data  $y_i$ , for  $i = 0, \dots, 179$ , adding a gaussian distributed noise of zero mean and variance  $\sigma^2 = 100$ . The noise in the synthetic data is shown in Figure 16 b) in aggregate form. For the discrepancy principle (2.2) we choose  $\tau = 2$ . See Table 7. We observe, from Figure 16 and Figure 17, how the presence of  $A$  (and the fact that the information of the ‘‘true image’’ are included in  $A$ ) ‘‘helps’’ the iterates (3.18) in a better reconstruction of the value of the pressure than the Landweber iterates (3.19). However, in Figure 16 c) we can observe some evident artifacts which come from the other a-priori information contained in  $A$ .

**Test 2aS/2bS.** (Figures 18 and 20). *Target: reconstruct one of the image which has been utilized to create the matrix  $A$ , with regions with only negative values.* See Figure 18 for the ‘‘True Image’’. All the synthetic data  $y_i^\delta$ , for  $i = 0, \dots, 179$ , are obtained by the real data  $y_i$ , for  $i = 0, \dots, 179$ , adding a gaussian distributed noise of zero mean and variance  $\sigma^2 = 100$  in Test 2aS, and  $\sigma^2 = 10$  in Test 2bS. The noise in the synthetic data for the two tests is shown in Figure 18 b) and Figure 20 in aggregate form. For the discrepancy principle (2.2) we choose  $\tau = 3$  in Test 2aS and  $\tau = 6$  in Test 2bS. See Table 8 and Table 9. Again, in Test 2aS we notice how the presence of  $A$  ‘‘helps’’ the iterates (3.18) to get a better reconstruction of all the negative values of the pressure than the Landweber iterates (3.19), see Figure 18 and Figure 19. However, we also observe some evident artifacts which come from of the other a-priori in  $A$ . In Test 2bS, we consider the case of limited data, which corresponds in choosing the data only of

some directions in  $\Theta$  (see (a) for its definition) and set to zero the values of the other vectors related to the angles not chosen, see Figure 6 b) for the directions of the available data. Comparing the results of the two schemes, (3.18) and (3.19), we can observe a good reconstruction for the iteratively regularized Landweber iteration.

**Test 3S.** (Figure 22). *Target: reconstruct one of the image which has been used to create the matrix  $A$ , with regions with positive or negative values.* See Figure 22 for the true image. All the synthetic data  $y_i^\delta$ , for  $i = 0, \dots, 179$ , are obtained by the real data  $y_i$ , for  $i = 0, \dots, 179$ , adding a gaussian distributed noise of zero mean and variance  $\sigma^2 = 100$ . For the discrepancy principle (2.2) we choose  $\tau = 3$ . See Table 10.

We stress that Figures 17 c), 19 c), 21 c), 23 c) show the ratios between the residuals of the data driven model and the model driven approach. A value greater than 1 of the ratio implies that the data driven approach has a significant influence on the iteration process.

As in the previous section, Figures 17 C), 19 C), 21 C) and 23 C) show the ratios between the residuals of the data driven model,  $\|Au_k - y^\delta\|$ , and the model driven approach,  $\|F(u_k) - y^\delta\|$ . A value greater than 1 of the ratio implies that the data driven approach has a significant influence on the iteration process. Figures 17 D), 19 D), 21 D) and 23 D) enlight how fast the product  $\|Ru_k - y^\delta\|_2 \|Au_k - y^\delta\|_2$  is going to zero.

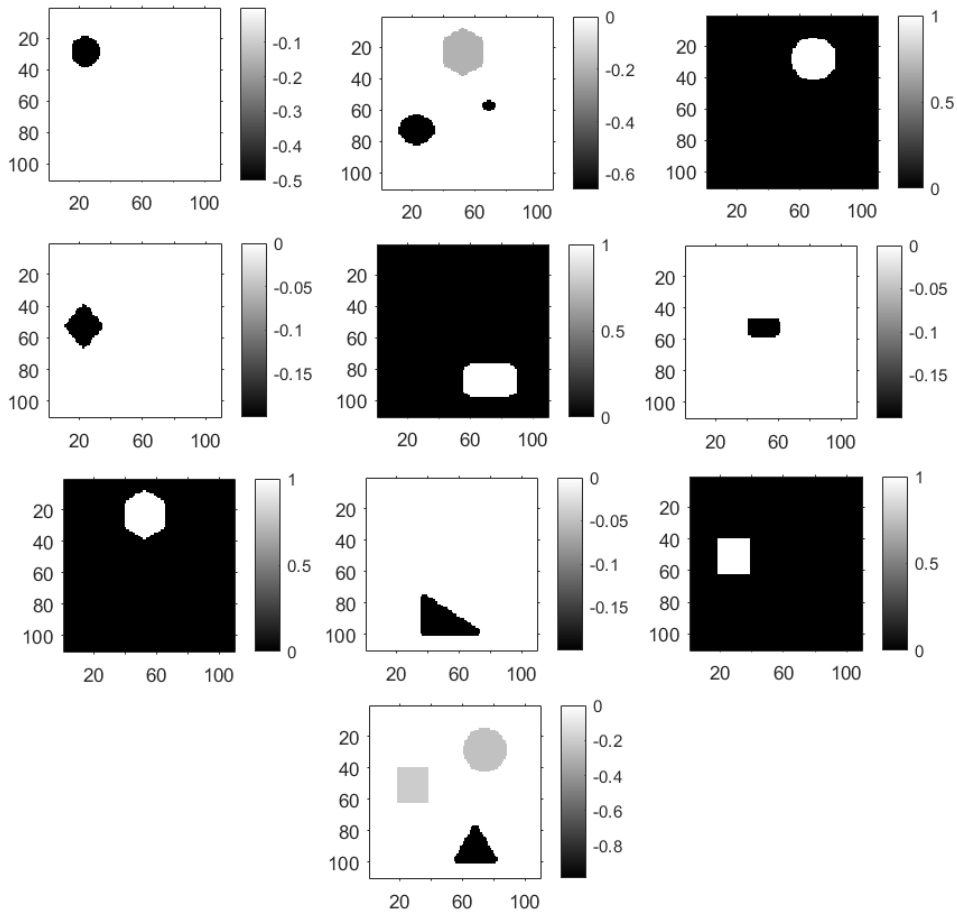


FIGURE 14. Dataset utilized to create the matrix  $A$  in Tests 1S, 2aS and 2bS.

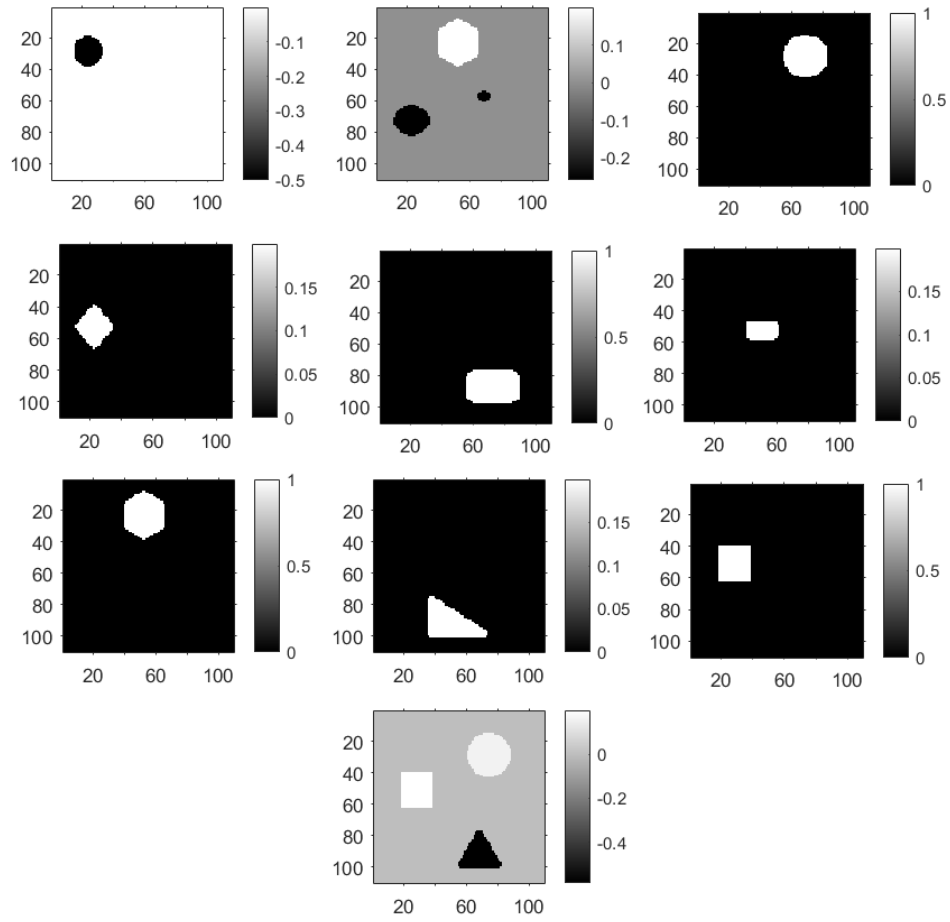
FIGURE 15. Dataset utilized to create the matrix  $A$  in the following Test 3S.

TABLE 7. Test 1S. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	100	141,519952	2	73	34,9193152	0,224935913
IRLI				128	7,5506274	0,233332376
LANDWEBER				190	11,3537075	0,247679701

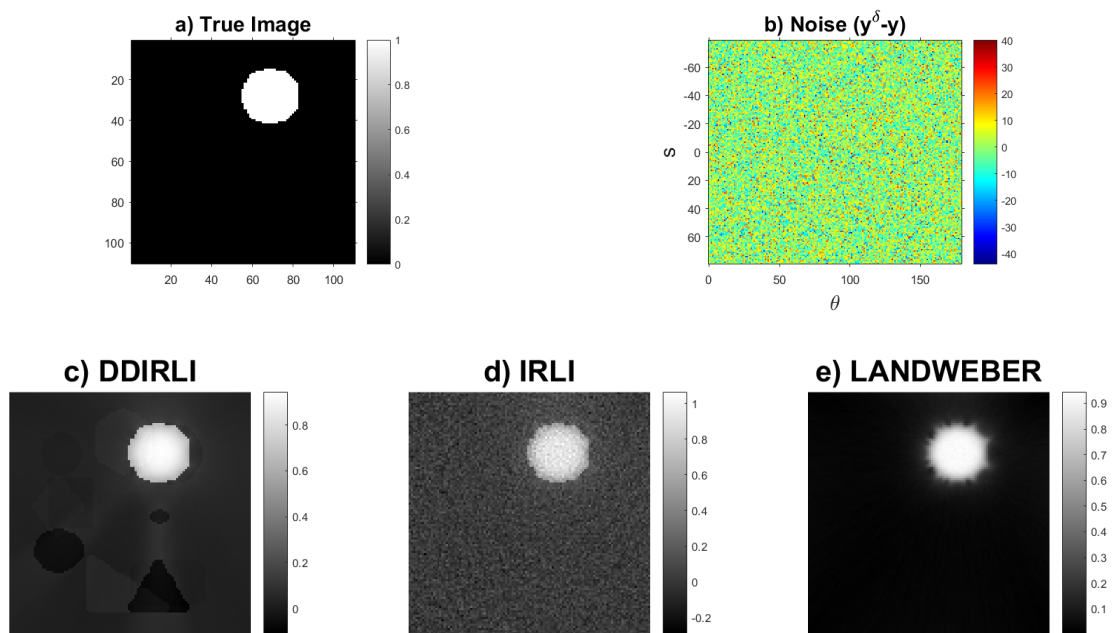


FIGURE 16. Test 1S. Reconstruction of an image which has been used to build the matrix  $A$ , only with a region with a positive value. a) Image to be reconstructed; b) Plot of the noise added in the synthetic data, which are aggregated in a matrix form; c)-e) Reconstructions by the different methods.

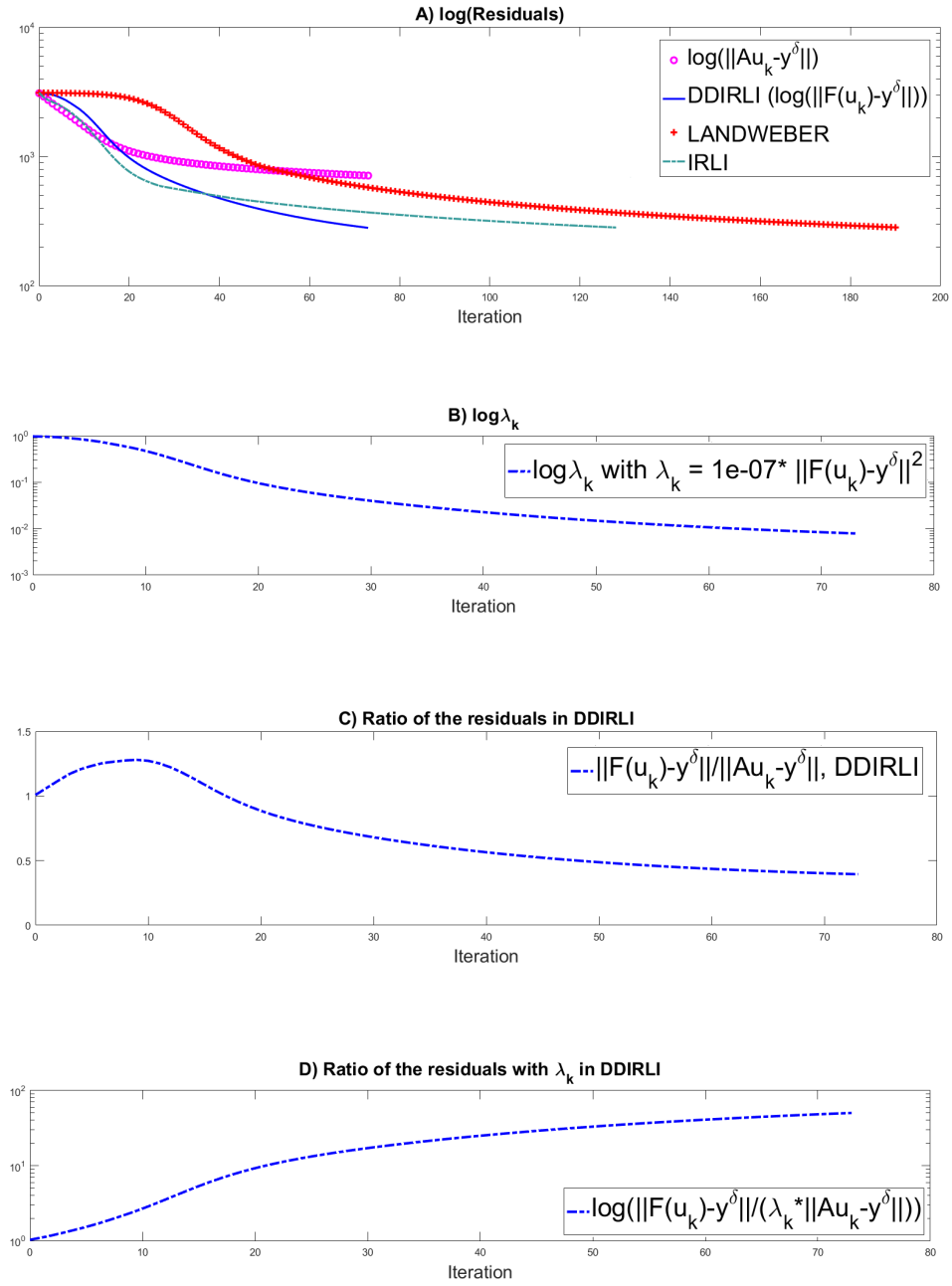


FIGURE 17. Test 1S. A) Residual at each step of schemes (3.18), (3.19) and (3.20); B) values of  $\lambda_k$  in (3.18) scheme at each iteration; C) ratio of the residuals  $\|F(u_k) - y^\delta\|_\infty / \|Au_k - y^\delta\|_\infty$  at each iteration; D) plot of the ratio  $\|F(u_k) - y^\delta\|_\infty / (\lambda_k \|Au_k - y^\delta\|_\infty)$ .

TABLE 8. Test 2aS. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	100	144,6648313	3	42	20,0855993	0,18951741
IRLI				66	4,3175441	0,30307348
LANDWEBER				156	9,0376626	1,802786044

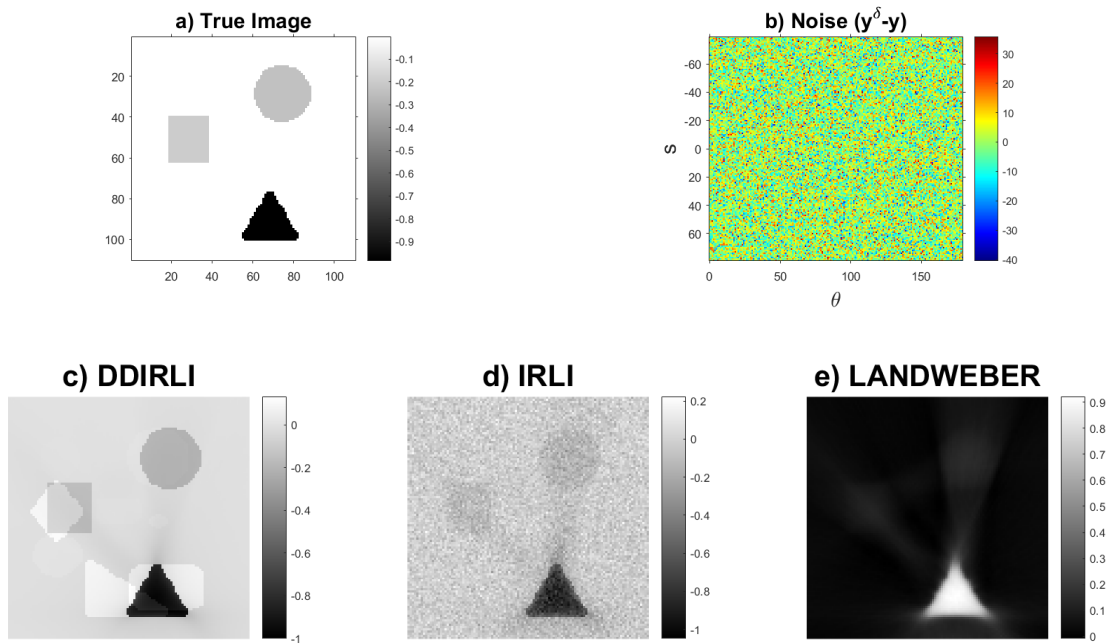


FIGURE 18. Test 2aS. Reconstruction of an image which has been used to build the matrix  $A$ , only with regions with negative values. a) Image to be reconstructed; b) Plot of the noise added in the synthetic data, which are aggregated in a matrix form; c)-e) Reconstructions by the different methods.

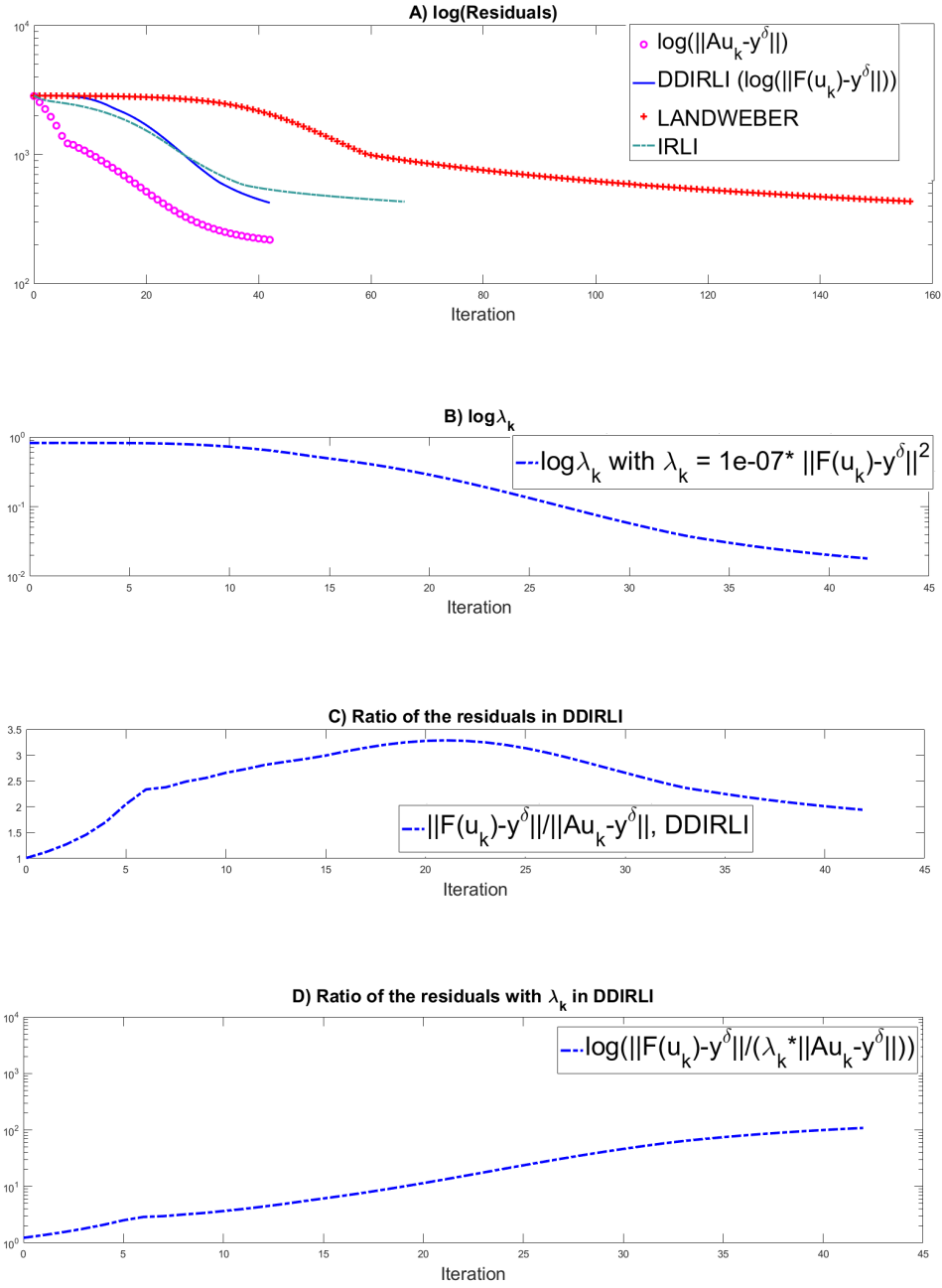


FIGURE 19. Test 2aS. A) Residual at each step of schemes (3.18), (3.19) and (3.20); B) values of  $\lambda_k$  in (3.18) scheme at each iteration; C) ratio of the residuals  $\|F(u_k) - y^\delta\|_\infty / \|Au_k - y^\delta\|_\infty$  at each iteration; D) plot of the ratio  $\|F(u_k) - y^\delta\|_\infty / (\lambda_k \|Au_k - y^\delta\|_\infty)$ .

TABLE 9. Test 2bS. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	10	44,46384766	6	84	37,8148317	0,144113319
IRLI				399	24,3457831	0,262173351
LANDWEBER				399	24,076839	1,77556498

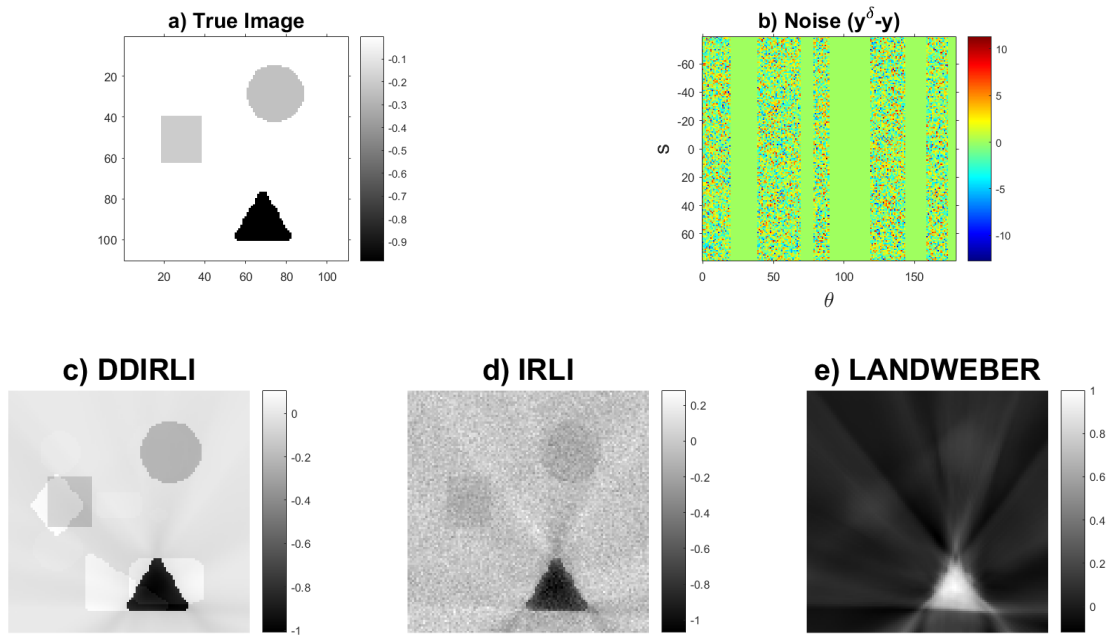


FIGURE 20. Test 2bS. Reconstruction of an image which has been used to build the matrix  $A$ , from partial data and only with regions with negative values. a) Image to be reconstructed; b) Plot of the noise added in the synthetic data, which are aggregated in a matrix form; c)-e) Reconstructions by the different methods.



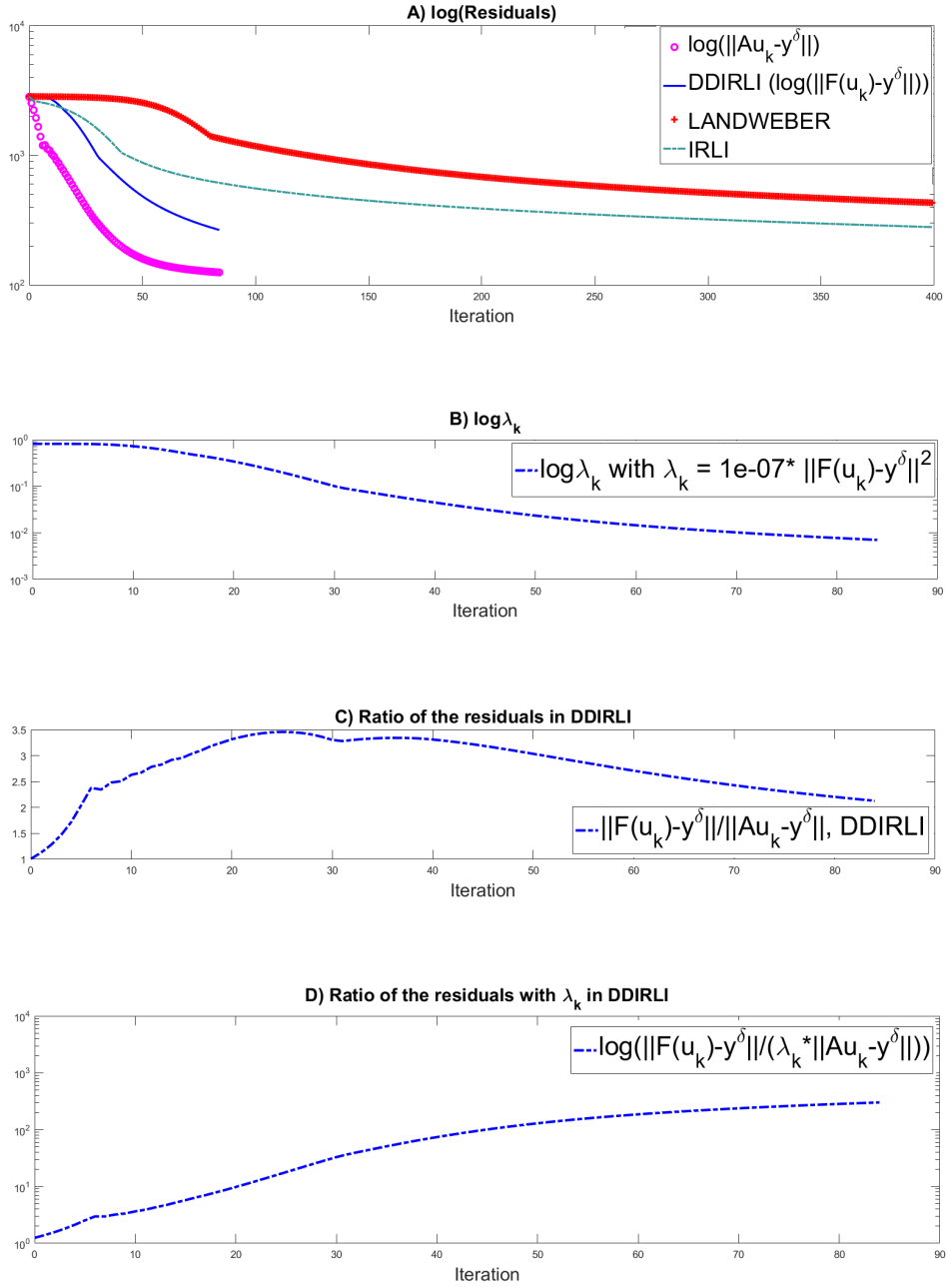


FIGURE 21. Test 2bS. A) Residual at each step of schemes (3.18), (3.19) and (3.20); B) values of  $\lambda_k$  in (3.18) scheme at each iteration; C) ratio of the residuals  $\|F(u_k) - y^\delta\|_\infty / \|Au_k - y^\delta\|_\infty$  at each iteration; D) plot of the ratio  $\|F(u_k) - y^\delta\|_\infty / (\lambda_k \|Au_k - y^\delta\|_\infty)$ .

TABLE 10. Test 3S. Left part: Parameters used in the test. Right part: some of the results of the test.

Method	$\sigma^2$ -noise	$\delta$	$\tau$	Iterations	Comp. Time (s)	$\frac{\ u_{\text{true}} - u_{\text{rec}}\ _2}{\ u_{\text{true}}\ _2}$
DDIRLI	100	145,8908582	3	213	94,3754666	0,254685973
IRLI				83	4,7161903	0,475122924
LANDWEBER				151	8,5576423	1,372143804

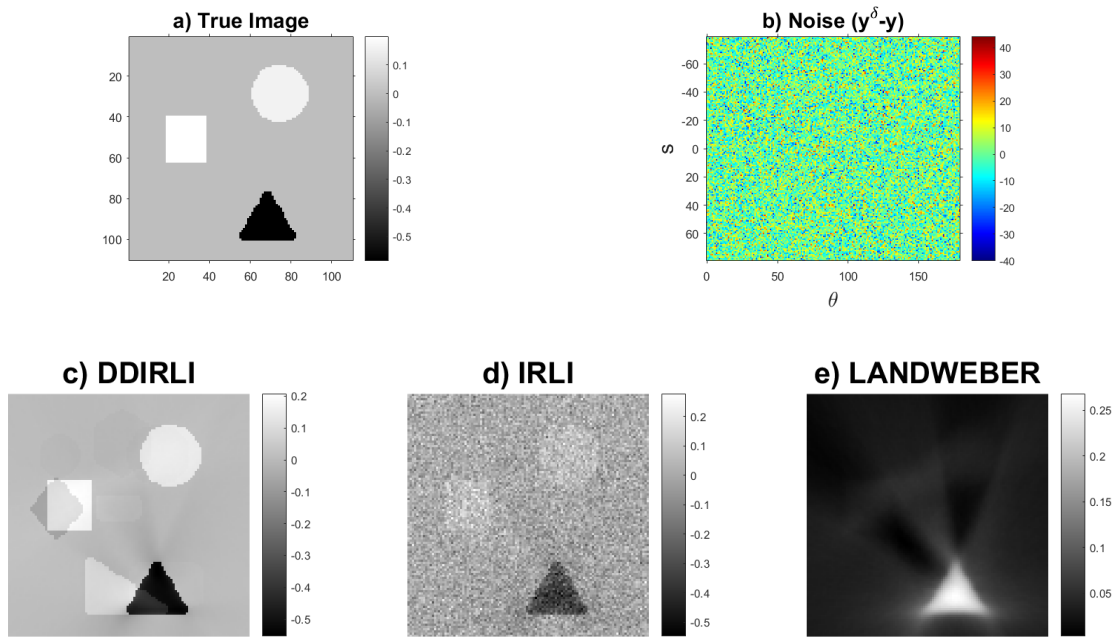


FIGURE 22. Test 3S. Reconstruction of an image which has been used to build the matrix  $A$ , with regions with positive and negative values. a) Image to be reconstructed; b) Plot of the noise added in the synthetic data, which are aggregated in a matrix form; c)-e) Reconstructions by the different methods.

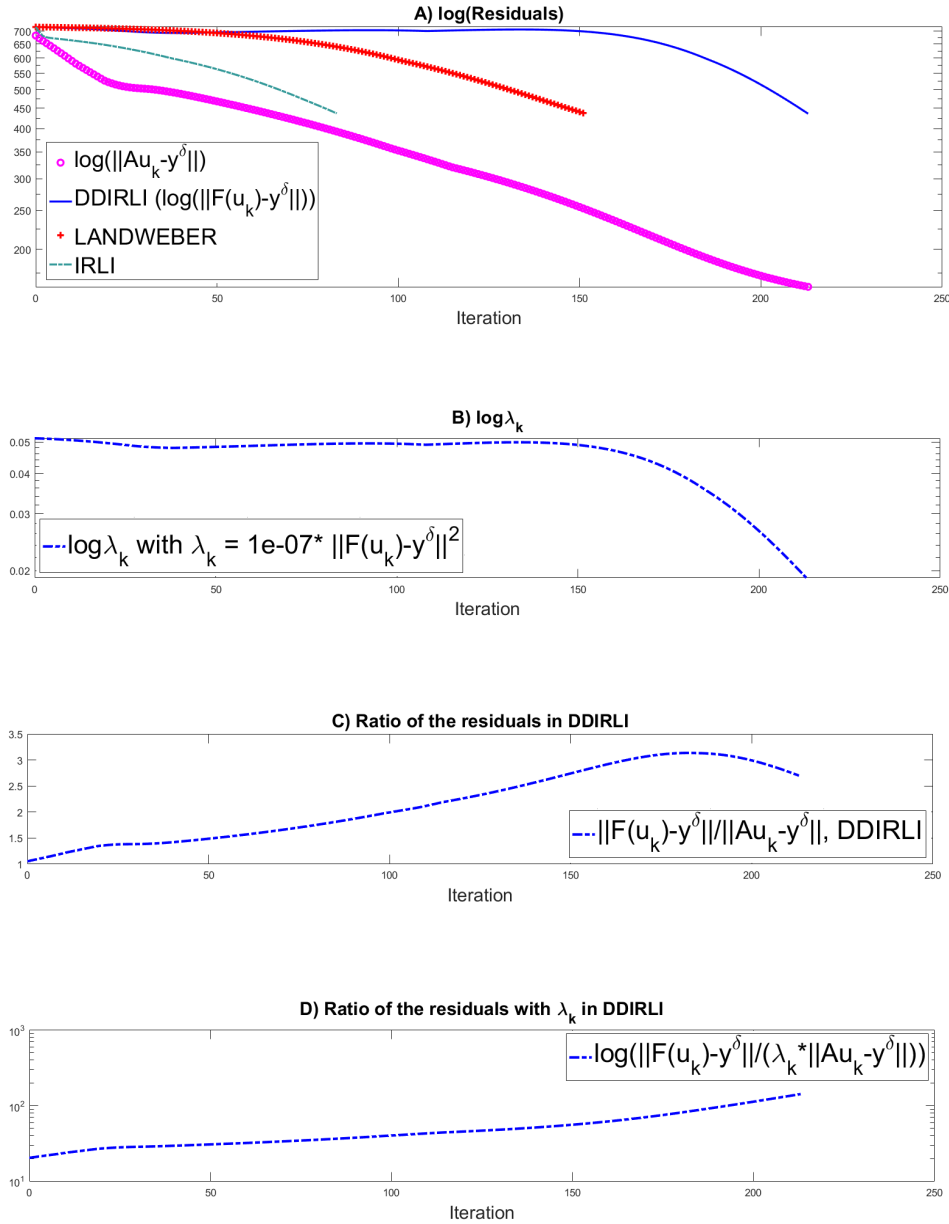


FIGURE 23. Test 3S. A) Residual at each step of schemes (3.18), (3.19) and (3.20); B) values of  $\lambda_k$  in (3.18) scheme at each iteration; C) ratio of the residuals  $\|F(u_k) - y^\delta\|_\infty / \|Au_k - y^\delta\|_\infty$  at each iteration; D) plot of the ratio  $\|F(u_k) - y^\delta\|_\infty / (\lambda_k \|Au_k - y^\delta\|_\infty)$ .

## 4. CONCLUSION

In this paper we introduced a new variant of the iteratively regularized Landweber iteration, where additional regularization is enforced from training data. The data driven regularization term is determined by finding a linear system from the input-output relation of some training data.

This strategy is uneconomical in terms of the amount of usable training data, stability and matrix storage capacity, and this asks for more advanced methods of learning (see for example [1, 7]) instead of a linear black-box strategy. For some literature on deep learning in inverse problems, see for example [1, 2, 4, 6] and references therein. We believe that the issue, of the use of a large amount of storage capacity for the creation of the matrix A, could be overcome by using other methods of numerical linear algebra which are equally valid and precise compared to the singular value decomposition. We leave this issue for future

works. Looking closer to our proposed algorithm it averages data driven regularization and physical model terms in such a way that in the beginning of the iteration the data driven term is dominant, while during the iteration the physical model takes over the leading role. By this sense we make sure that for data with little noise we are always close to the physical solution. Our numerical experiments show that the iteration always reconstructs an image where the residuum is close to the measured data. Moreover, if the desired solution is well approximated by the training data, the reconstruction with the proposed algorithm always works better than the standard Landweber iteration and the iteratively regularized Landweber iteration. This, in particular, applies to severely ill-posed problems of limited angle tomography. A significant challenge are nonlinear inverse problems, such as the problem of Schlieren tomography. The inherent non-uniqueness of the solution of the Schlieren problem gives more emphasis on selecting the right a-priori choice, which our flexible regularization term can handle more efficiently than existing theory. In the nonlinear case, one can observe various artifacts in the reconstructions due to the linearity of  $A$ , which are some of the images contained in  $A$ . Therefore, the use of a nonlinear operator  $A$  should be investigated. We leave this issue for future works.

### ACKNOWLEDGMENTS

A. Aspri thanks Y. Korolev for suggesting the dataset MNIST, used in the first part of the paper. O. Scherzer is supported by the FWF via the projects I3661-N27 (Novel Error Measures and Source Conditions of Regularization Methods for Inverse Problems) and via SFB F68, project F6807-N36 (Tomography with Uncertainties).

### REFERENCES

- [1] J. Adler and O. Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):124007, 24, 2017.
- [2] S. Antholzer, M. Haltmeier, and J. Schwab. Deep learning for photoacoustic tomography from sparse data. *Inverse Probl. Sci. Eng.*, 27(7):987–1005, 2019.
- [3] A. B. Bakushinskii. On a convergence problem of the iterative-regularized Gauss-Newton method. *Zh. Vychisl. Mat. i Mat. Fiz.*, 32(9):1503–1509, 1992.
- [4] T. A. Bubba, G. Kutyniok, M. Lassas, M. März, W. Samek, S. Siltanen, and V. Srinivasan. Learning the invisible: a hybrid deep learning–shearlet framework for limited angle computed tomography. *Inverse Problems*, 35(6):064002, 38, 2019.
- [5] T. Charlebois and R. Pelton. Quantitative 2d and 3d schlieren imaging for acoustic power and intensity measurements. *Medical Electronics*, pages 789–792, 1995.
- [6] F. Dreier, S. Pereverzyev, Jr., and M. Haltmeier. Operator learning approach for the limited view problem in photoacoustic tomography. *Comput. Methods Appl. Math.*, 19(4):749–764, 2019.
- [7] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2016.
- [8] M. Haltmeier, R. Kowar, A. Leitão, and O. Scherzer. Kaczmarz methods for regularizing nonlinear ill-posed equations. II. Applications. *Inverse Probl. Imaging*, 1(3):507–523, 2007.
- [9] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.
- [10] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.
- [11] R. Kowar. Reconstruction of transducer pressure fields from Schlieren data. In *Progress in industrial mathematics at ECMI 2006*, volume 12 of *Math. Ind.*, pages 548–552. Springer, Berlin, 2008.

- 
- [12] P. Kuchment. *The Radon transform and medical imaging*, volume 85 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.
  - [13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
  - [14] G. Ólafsson and E. T. Quinto. *The Radon Transform, Inverse Problems, and Tomography*. Number 63 in *Proceeding of Symposia in Applied Mathematics*. American Mathematical Society, Providence, Rhode Island, 2006.
  - [15] A. Papoulis. *The Fourier integral and its applications*. McGraw-Hill Book Co. Inc. New York-San Francisco, Calif. London Toronto, 1962.
  - [16] T. A. Pitts, J. F. Greenleaf, J. L. Jian-Yu Lu, and R. R. Kinnick. Tomographic schlieren imaging for measurement of beam pressure and intensity. In *1994 Proceedings of IEEE Ultrasonics Symposium*, volume 3, pages 1665–1668, 1994.
  - [17] O. Scherzer. A modified Landweber iteration for solving parameter estimation problems. *Appl. Math. Optim.*, 38(1):45–68, 1998.
  - [18] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
  - [19] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2012.