A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems

R. Ramlau, G. Teschke, M. Zhariy

October 8, 2008

Abstract

In this paper we shall be concerned with the construction of an adaptive Landweber iteration for solving linear ill-posed and inverse problems. Classical Landweber iteration schemes provide in combination with suitable regularization parameter rules order optimal regularization schemes. However, for many applications the implementation of Landweber's method is numerically very intensive. Therefore we propose an adaptive variant of Landweber's iteration that significantly may reduce the computational expense, i.e. leading to a compressed version of Landweber's iteration. We lend the concept of adaptivity that was primarily developed for well-posed operator equations (in particular, for elliptic PDE's) essentially exploiting the concept of wavelets (frames), Besov regularity, best N-term approximation and combine it with classical iterative regularization schemes.

As the main result of this paper we define an adaptive variant of Landweber's iteration. In combination with an adequate refinement/stopping rule (a-priori as well as a-posteriori principles) we prove that the proposed procedure is a regularization method which converges in norm for exact and noisy data. The proposed approach is verified in the field of computerized tomography imaging.

MSC: 15A29, 47A52, 68U10, 94A08, 94A40

Keywords: Compressive algorithms, linear ill-posed inverse problem, nonlinear and adaptive approximation, sparsity

1 Introduction

We address the problem of computing an approximation to a solution of a linear operator equation $Lf = g$, where $L : \mathcal{H} \to \mathcal{H}'$ is a linear operator between Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ and where (in many relevant cases) only noisy data g^δ with $\|g^\delta - g\| \leq \delta$ are available. In this framework we are often faced with the ill-posedness of the operator equation and therefore with regularization issues.

In the last decades, several regularization methods have been established for linear as well as nonlinear inverse problems. In principle, there exist regularization methods that are based on Tikhonov's approach (i.e. adding regularizing constraints), iteration methods and discretization approaches. For a comprehensive discussion on this subject we refer the reader to the rich literature, see e.g., [17, 20]. For each of those methods abundant refinements/generalizations and different regularization parameter rules were established, e.g. the discrepancy principle [22, 33, 19], the monotone error rule [31], or, more recently, the Lepskii principle [2, 21], just to name a few. The convergence of any of the investigated methods towards the solution of the equation can be in general arbitrary slow. In order to obtain convergence rates, smoothness or range conditions are commonly used. However, within the last 5 years it has been discovered that a sparsity assumption also leads to highly stable algorithms. A sparse reconstruction, i.e. a reconstruction that has only few non-zero frame coefficients, is usually obtained by adding a constraint that penalizes non-sparse expansions. One of the first papers in which a comprehensive and coherent analysis on linear inverse problems and sparsity was given is [10]. In this paper, the authors consider the Tikhonov approach in which a sparsity constraints is involved and construct, based on Gaussian surrogate extension methods, some sort of projected Landweber iteration. Several generalizations of this approach, even to nonlinear problems, can be found in [1, 4, 12, 13, 14, 15, 28, 32]. If a proper basis or frame is chosen, then the use of the sparsity constraint allows to capture local morphological properties of the solution, e.g. discontinuities, which are represented with only few coefficients leading to very efficient expansions. In the past this could only be achieved at a high computational complexity combined with a slow convergence speed of the employed iteration method. Currently, there are several accelerated iteration methods (or so-called compressed algorithms) on the market that yield for instance through domain decomposition strategies, see [18], or through projection strategies, see [11], with much less iterations reasonable results. These methods, however, act up to the domain decomposition or projection step on all involved frame functions in a uniform way (even on those atoms that are absolutely not involved in representing the solution). One way to circumvent this unprofitable proceeding is to operate in an adaptive way, i.e. operate only with those frame coefficients or matrix entries that really contribute to a good approximation of the solution. Such adaptive strategies that bound the computational complexity and ensure an optimal approximation order have been proposed in the past e.g. in [6, 29]. In this context, adaptivity is based on a well selected discretization of the operator leading to compressible matrices that allow efficient matrix vector multiplications. So far, these methods have been only used for well-posed operator equations [29] or already regularized/stabilized ill-posed problems [3, 8].

To overcome this shortfall and provide adaptive techniques also for more general ill-posed problems is the goal of the present paper. The analysis presented here essentially combines building blocks from the theory of linear inverse problems and the theory of adaptive methods for well-posed problems.

Main Result. As the main result we provide an adaptive Landweber iteration scheme to approximate the solution of the inverse problem $Lf = g^\delta$ for which we show norm convergence for exact and noisy data and regularization properties for a-priori and a-posteriori parameter selection rules.

Organization of the Paper. The remaining paper is organized as follows: In Section 2, we review basic ideas and results on adaptive approximation, ill-posed inverse problems and iterative regularization techniques. The main results are given in Section 3 and 4 in which we show that an adaptive evaluation of Landweber's iteration is indeed a regularization scheme (for a-priori as well as for a-posteriori parameter rules). Finally, in Section 5 we present a first numerical experiment in the context of computerized tomography.

2 Adaptivity, inverse problems and Landweber's iteration

Within this section we briefly recall the setup of inverse problems as it is required for later analysis. We are in particular interested in Landweber's iteration which is known to be an order optimal method both for a-priori or a-posteriori regularization parameter rules. As mentioned above, our main focus will be on the development of a nonlinear and adaptive version of the Landweber iteration, which requires an appropriate discretization of the operator equation $Lf = g$. Here, we will use a discretization generated by a suitable wavelet basis or frame which will allow us to treat the operator as a compressible infinite matrix (see Section 2.2).

The reason why to apply nonlinear and adaptive strategies is the expected performance of the iteration. In principle, adaptive strategies perform better than standard linear methods only if certain properties are fulfilled. To clarify this statement, we introduce by f_N the best N-term approximation for f, i.e., a vector with at most N nonzero coefficients that has distance to f less than or equal to that of any vector with a support of that size. Considering bases or frames of sufficiently smooth wavelet type (e.g. wavelets of order n), it is shown in [6, 29] that if both

$$0 < s < \frac{d-t}{n} ,$$

where n is the space dimension and t denoting the the smoothness of the Sobolev space where L has its domain, and f is in the Besov space $B_\tau^{sn+t}(L_\tau)$ with $\tau = (1/2+s)^{-1}$, then

$$\sup_{N\in\mathbb{N}} N^s \|f - f_N\| < \infty .$$

The condition here involving Besov regularity is much milder than requiring $f \in H^{sn+t}$ that would be needed to guarantee the same rate of convergence with linear approximation. Indeed, it has been verified for some boundary value problems that the Besov regularity is much higher than the Sobolev regularity, and in these cases adaptivity pays off. However, for inverse problems it is in general not always possible to estimate the regularity of the solution from the regularity of the right hand side due to the presence of the noise. Typically, convergence rates and optimality are obtained by assuming so-called source conditions, which are based on specific a-priori knowledge about the solution and/or the operator. In certain cases, the source condition directly translates into smoothness in scales of Sobolev spaces. As pointed out in [23], for the particular tomographic reconstruction problem, a suitable model class are piecewise constant functions with jumps along smooth manifolds. It is shown that such functions belong to the Sobolev space H^{sd} with $sd < 1/2$. An adaptive approximation of such functions (when carried out in L_2) pays off if the Besov regularity in the scale $B_\tau^{sd}(L_\tau(\Omega))$, $\tau = (s+1/2)^{-1}$ is significantly higher. This issue is discussed in [29, Rem. 4.3] and [34] and indeed such functions belong to $B_\tau^{sd}(L_\tau(\Omega))$ with $sd < 1/\tau = s + 1/2$. For the two-dimensional case, which is the case in our application, we therefore have that the solution to be reconstructed belongs to $H^{sd}(\Omega)$ for $s < 1/4$ and to $B_\tau^{sd}(L_\tau(\Omega))$ for $s < 1/2$. Consequently, the Besov regularity is indeed higher than the Sobolev regularity. Hence, the application of adaptive schemes to solve the tomographic reconstruction problem should be numerically more efficient.

However, for other frameworks/inverse problems no general results are available so far. Nevertheless, numerical experiments indicate that even then it seems to be very reasonable and worthy to apply the concept of best N-term approximations and associated numerical adaptive schemes.

2.1 Brief review on inverse problems and regularization results

In inverse problems, the typical goal is to compute an approximation to a solution of the ill-posed problem

$$Lf = g \tag{1}$$

from noisy data g^δ with $\|g - g^\delta\| \leq \delta$. For $g \in Ran(L) \oplus Ran(L)^\perp$ we apply the concept of a generalized solution $f^\dagger = L^\dagger g$, which is defined as the minimizer of $\|Lf - g\|$ with minimal norm, or equivalently, as the minimum norm solution of the normal equation

$$L^*Lf = L^*g . \tag{2}$$

For ill-posed operators one typically has $Ran(L) \neq \overline{Ran(L)}$, and the generalized inverse $L^\dagger$ is unbounded. In the presence of noisy data we therefore have to use regularization methods in order to compute a stable approximation to the solution. A family of continuous operators $\{T_\alpha\}_{\alpha\in\mathbb{R}^+}$ with a parameter choice rule $\alpha = \alpha(\delta)$ with $\lim_{\delta\to 0} \alpha(\delta) = 0$ is called a regularization for L, if

$$\lim_{\delta\to 0} T_{\alpha(\delta)} g^\delta = f^\dagger . \tag{3}$$

Well known examples for regularization methods are Tikhonov regularizations or iterative methods (e.g. the Landweber method). For a detailed investigation of regularization methods we refer to the textbooks [17, 20]. It is a well known result that the convergence in (3) can be arbitrarily slow. In order to derive convergence rate results, we need additional information on the searched for solution $f^\dagger$. In particular, we will assume that $f^\dagger \in X_\nu = \{f \in X : f = (L^*L)^{\nu/2}\omega\}$ with $\|f^\dagger\|_\nu := \|(L^*L)^{-\nu/2}f^\dagger\| = \rho < \infty$. Then a regularization method with parameter choice rule $\alpha(\delta)$ is called of optimal order if

$$\|T_{\alpha(\delta)}g^\delta - f^\dagger\| \le c_0\delta^{\nu/(\nu+1)}\rho^{1/(\nu+1)} \ . \tag{4}$$

Again we refer to [17, 20] for more details. In this paper, we will in particular focus on the Landweber iteration, which is a gradient method for the minimization of the residual

$$\|g^\delta - Lf\|^2 \ , \tag{5}$$

i.e. the iteration is defined by

$$f^\delta_{m+1} = f^\delta_m - \beta L^*(Lf^\delta_m - g^\delta) \ . \tag{6}$$

As it can be retrieved, e.g. in [20], iteration (6) is for $0 < \beta < 2/\|L\|^2$ a linear regularization method for problem (1) as long as the iteration is truncated at some finite index m_*. In order to identify the optimal truncation index m_*, one may apply either an a-priori or an a-posteriori parameter rule. The Landweber method (6) is an order optimal linear regularization method [20] if the iteration is truncated at the a-priori chosen iteration index

$$m_* = \lfloor \beta\big(2\frac{\beta}{\nu}e\big)^{\nu/(\nu+1)}\big(\frac{\rho}{\delta}\big)^{2/(\nu+1)}\rfloor \ , \tag{7}$$

where the common notation $\lfloor x \rfloor$ denotes the smallest integer less or equal x. If m_* is chosen as suggested in (7), then optimal convergence order with respect to $f^\dagger$ can be achieved with the constant c_0 in (4) given by $c_0 = (1+\nu)(2\nu e)^{-\nu(2\nu+2)}$. This proceeding, however, needs exact knowledge of the parameters ν, ρ in the source condition. This shortfall can be avoided when applying Morozov's discrepancy principle. This principle performs the iteration as long as

$$\|Lf^\delta_m - g^\delta\| > \tau\delta \tag{8}$$

holds with $\tau > 1$, and truncates the iteration once

$$\|Lf^\delta_{m_*} - g^\delta\| \le \tau\delta \tag{9}$$

is fulfilled for the first time. The regularization properties of this principle were investigated in [16]. The authors have shown that, as long as (8) holds, the next iterate will be closer to the generalized solution than the previous iterate. This property turned out to be very fruitful for the investigation of discretized variants of (6). This can be retracted in greater detail [25] in which a discretization of the form

$$f^\delta_{m+1} = f^\delta_m - \beta L^*_{r^\delta(m)}(L_{r^\delta(m)}f^\delta_m - g^\delta) \tag{10}$$

was suggested. The basic idea in [25] is the introduction of approximations $L_{r^\delta(m)}$ to the operator L that are updated/refined in dependance on a specific discrepancy principle.

2.2 Discretization and the adaptive evaluation of Landweber's method

Let the space X be a separable Hilbert space and assume we are given a preassigned system of functions $\{\phi_\lambda : \lambda \in \Lambda\} \subset X$ for which there exists constants A, B with $0 < A \le B < \infty$ such that for all $x \in X$,

$$A\|x\|^2_X \le \sum_{\lambda\in\Lambda} |\langle x, \phi_\lambda\rangle|^2 \le B\|x\|^2_X. \tag{11}$$

For such a system, which is called a frame for X (ensuring more flexibility than bases), see [5], we may consider the operator

$$F : X \to \ell_2 \text{ via } x \mapsto \boldsymbol{x} = \{\langle x, \phi_\lambda \rangle\}_{\lambda \in \Lambda}$$

with adjoint

$$F^* : \ell_2 \to X \text{ via } \boldsymbol{f} \mapsto \sum_{\lambda \in \Lambda} \boldsymbol{f}_\lambda \phi_\lambda .$$

The operator F is often referred in the literature to as the analysis operator, whereas F^* is referred to as the synthesis operator. The composition of both, F^*F, is called the frame operator which is by condition (11) an invertible map; guaranteing that each $x \in X$ can be reconstructed from its moments $\langle x, \phi_\lambda \rangle$. Moreover, for the sequence space ℓ_2 one has $\ell_2 = \text{Ran } F \oplus \text{Ker } F^*$. In order to create a discrete version of iteration (6), we discretize normal equation (2) by defining

$$\boldsymbol{S} = FL^*LF^* \ , \quad f = F^*\boldsymbol{f} \quad \text{and} \quad \boldsymbol{g}^\delta = FL^*g^\delta$$

and obtain an equivalent ℓ_2 problem,

$$\boldsymbol{S}\boldsymbol{f} = \boldsymbol{g}^\delta \ . \tag{12}$$

An approximate solution for (12) can be derived by the corresponding sequence space Landweber iteration,

$$\boldsymbol{f}^\delta_{m+1} = \boldsymbol{f}^\delta_m - \beta(\boldsymbol{S}\boldsymbol{f}^\delta_m - \boldsymbol{g}^\delta) \ . \tag{13}$$

Note that the operator $\boldsymbol{S} : \ell_2 \to \ell_2$ is symmetric but through the ill-posedness of L not boundedly invertible on ℓ_2 (even on the subspace Ran F). This is the major difference to [29] in which the invertibility of $\boldsymbol{S}$ on Ran F was substantially used to ensure the convergence of the damped Richardson iteration (coinciding with Landweber iteration (13) in our setup).

Since in actual computations we can neither handle the infinite dimensional vectors $\boldsymbol{f}^\delta_m$ and $\boldsymbol{g}^\delta$ nor apply the infinite dimensional matrix $\boldsymbol{S}$, iteration (13) is not a practical algorithm. To this end, we need to study the convergence and regularization properties of the iteration in which $\boldsymbol{f}^\delta_m$, $\boldsymbol{g}^\delta$ and $\boldsymbol{S}$ are approximated by finite length objects. Proceeding as suggested [29], we assume that we have the following three routines at our disposal:

- $\boldsymbol{RHS}_\varepsilon[g] \to \boldsymbol{g}_\varepsilon$. This routine determines a finitely supported $\boldsymbol{g}_\varepsilon \in \ell_2$ satisfying

$$\|\boldsymbol{g}_\varepsilon - FL^*g\| \le \varepsilon \ .$$

- $\boldsymbol{APPLY}_\varepsilon[\boldsymbol{f}] \to \boldsymbol{w}_\varepsilon$. This routine determines, for a finitely supported $\boldsymbol{f} \in \ell_2$ and an infinite matrix $\boldsymbol{S}$, a finitely supported $\boldsymbol{w}_\varepsilon$ satisfying

$$\|\boldsymbol{w}_\varepsilon - \boldsymbol{S}\boldsymbol{f}\| \le \varepsilon \ .$$

- $\boldsymbol{COARSE}_\varepsilon[\boldsymbol{f}] \to \boldsymbol{f}_\varepsilon$. This routine creates, for a finitely supported with $\boldsymbol{f} \in \ell_2$, a vector $\boldsymbol{f}_\varepsilon$ by replacing all but N coefficients of $\boldsymbol{f}$ by zeros such that

$$\|\boldsymbol{f}_\varepsilon - \boldsymbol{f}\| \le \varepsilon \ ,$$

  whereas N is at most a constant multiple of the minimal value N for which the latter inequality holds true.

For the detailed functionality of these routines we refer the interested reader to [6, 29]. As the $\boldsymbol{COARSE}$ routine is of particular interest, a detailed description will be given below. For the sake of more flexibility in our proposed approach, we allow (in contrast to classical setup suggested in [29]) ε to be different within each iteration step and sometimes different for each of the three routines. Consequently, we set $\varepsilon = \varepsilon^R_m$ for the routine $\boldsymbol{RHS}_\varepsilon[\cdot]$, $\varepsilon = \varepsilon^A_m$ for $\boldsymbol{APPLY}_\varepsilon[\cdot]$ and, finally, $\varepsilon = \varepsilon^C_m$ for $\boldsymbol{COARSE}_\varepsilon[\cdot]$. The subscript m of the created error tolerance or so-called refinement sequences $\{\varepsilon_m^C\}_{m\in\mathbb{N}}$, $\{\varepsilon_m^A\}_{m\in\mathbb{N}}$ and $\{\varepsilon_m^R\}_{m\in\mathbb{N}}$ will be related to the iteration index by specific refinement strategies of the form

$$r^\delta : \mathbb{N} \to \mathbb{N}.$$

In principle, the refinement sequences are zero sequences and must be selected in advance; the map r^δ represents a specific integer to integer map (constructed below) that allows an adjustment of the reconstruction accuracy within each iteration step m. As a simple example consider the refinement rule $r^\delta(m) = m$ that chooses for each iteration m the preselected error tolerances ε_m^C, ε_m^A and ε_m^R. Choosing proper refinement strategies $r^\delta(m)$ enables us to establish convergence results and, thanks to the introduced subtleness, several desired regularization results.

For ease of notation we often write, if not otherwise stated, instead of $\varepsilon_{r^\delta(m)}^{\{C,A,R\}}$ just the index $r^\delta(m)$, i.e. we abbreviate $\boldsymbol{COARSE}_{\varepsilon^C_{r^\delta(m)}}[\cdot]$, $\boldsymbol{APPLY}_{\varepsilon^A_{r^\delta(m)}}[\cdot]$ and $\boldsymbol{RHS}_{\varepsilon^R_{r^\delta(m)}}[\cdot]$, respectively, by $\boldsymbol{COARSE}_{r^\delta(m)}[\cdot]$, $\boldsymbol{APPLY}_{r^\delta(m)}[\cdot]$ and $\boldsymbol{RHS}_{r^\delta(m)}[\cdot]$, respectively. Note, this does not mean the same accuracy for all three routines, it just means the same index for the accuracy/refinement sequences.

Summarizing our constructions, we define with the mentioned approximation procedures and the shorthand notations the following inexact/approximative variant of (13)

$$\tilde{\boldsymbol{f}}^\delta_{m+1} = \boldsymbol{COARSE}_{r^\delta(m)}\left[\tilde{\boldsymbol{f}}^\delta_m - \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\right] . \tag{14}$$

In what follows, we shall discuss in greater detail the regularization properties of the latter iteration for different regularization parameter rules.

3 Regularization theory for a-priori parameter rules

As reviewed in the previous section, the a-priori parameter rule (7) for the exact Landweber iteration (6) yields an order optimal regularization scheme. The natural question is whether the same holds true for the inexact (nonlinear and adaptive) Landweber iteration (14). A positive answer of the latter question essentially relies on the construction of a suitable refinement strategy r^δ.

In order to achieve an optimal convergence rate, we have to establish some preliminary results describing the difference between the exact iteration (6) and the inexact iteration (14).

Lemma 1. *Let $\boldsymbol{f}^\delta_m$ and $\tilde{\boldsymbol{f}}^\delta_m$ be defined as in* (6) *and* (14)*, respectively. Then, for all $m \geq 0$,*

$$\|\boldsymbol{f}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_{m+1}\| \leq (1+\beta\|\boldsymbol{S}\|)\|\boldsymbol{f}^\delta_m - \tilde{\boldsymbol{f}}^\delta_m\| + \varepsilon^C_{r^\delta(m)} + \beta(\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)}). \tag{15}$$

Proof. This result can be easily deduced, since for any $m \geq 0$ one has

$$\begin{aligned}
\|\tilde{\boldsymbol{f}}^\delta_{m+1} - \boldsymbol{f}^\delta_{m+1}\| &\leq \|\tilde{\boldsymbol{f}}^\delta_{m+1} - (\tilde{\boldsymbol{f}}^\delta_m - \beta\boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta\boldsymbol{RHS}_{r^\delta(m)}[g^\delta])\| \\
&\quad + \|\tilde{\boldsymbol{f}}^\delta_m - \beta\boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta\boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - (\boldsymbol{f}^\delta_m - \beta\boldsymbol{S}\boldsymbol{f}^\delta_m + \beta\boldsymbol{g}^\delta)\| \\
&\leq \varepsilon^C_{r^\delta(m)} + \|\tilde{\boldsymbol{f}}^\delta_m - \boldsymbol{f}^\delta_m - \beta\boldsymbol{S}\tilde{\boldsymbol{f}}^\delta_m + \beta\boldsymbol{S}\boldsymbol{f}^\delta_m\| + \beta\|\boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - \boldsymbol{g}^\delta\| \\
&\qquad + \beta\|\boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta\boldsymbol{S}\tilde{\boldsymbol{f}}^\delta_m\| \\
&\leq \varepsilon^C_{r^\delta(m)} + \beta(\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)}) + (1+\beta\|\boldsymbol{S}\|)\|\tilde{\boldsymbol{f}}^\delta_m - \boldsymbol{f}^\delta_m\| .
\end{aligned}$$

□

By a recursive application of Lemma 1, we achieve an explicit description of the difference.

Lemma 2. *Assume, $\boldsymbol{f}_0^\delta = \tilde{\boldsymbol{f}}_0^\delta$. Then, for all $m \geq 0$,*

$$\|\boldsymbol{f}_{m+1}^\delta - \tilde{\boldsymbol{f}}_{m+1}^\delta\| \leq \beta \sum_{i=0}^{m} (1 + \beta \|\boldsymbol{S}\|)^i (\varepsilon^C_{r^\delta(m-i)}/\beta + \varepsilon^A_{r^\delta(m-i)} + \varepsilon^R_{r^\delta(m-i)}). \tag{16}$$

Proof. Thanks to Lemma 1, we have for $m = 0$

$$\|\boldsymbol{f}_1^\delta - \tilde{\boldsymbol{f}}_1^\delta\| \leq \varepsilon^C_{r^\delta(0)} + \beta(\varepsilon^A_{r^\delta(0)} + \varepsilon^R_{r^\delta(0)}).$$

Now by induction for $m \to m+1$, we apply for $m \geq 1$ (15) and obtain (16). □

The latter lemmata allow now to prove that the truncated inexact Landweber iteration (14) is an order optimal regularization method. The regularization method T_α can be described with the help of an adequate refinement map r^δ and the a-priori parameter rule (7).

Definition 1 (Regularization method with a-priori parameter rule)**.**

*i) Given sequences of error tolerances $\{\varepsilon_m^{\{C,A,R\}}\}_{m\in\mathbb{N}}$ and routines **COARSE**, **APPLY** and **RHS** defined as above,*

ii) for $\delta > 0$ with $\|g^\delta - g\| \leq \delta$ derive the truncation index m_* as in (7)*,*

iii) define the quantities

$$C_{m,r^\delta} := \sum_{i=0}^{m} (1 + \beta\|\boldsymbol{S}\|)^i (\varepsilon^C_{r^\delta(m-i)} + \beta(\varepsilon^A_{r^\delta(m-i)} + \varepsilon^R_{r^\delta(m-i)})) \ ,$$

iv) choose the map r^δ such that C_{m_*-1,r^δ} satisfies

$$C_{m_*-1,r^\delta} \leq \delta^{\nu/(\nu+1)} \rho^{1/(\nu+1)} \ ,$$

v) define the regularization

$$T_\alpha g^\delta := F^* \tilde{\boldsymbol{f}}_{m_*}^\delta$$

with regularization parameter $\alpha = 1/m_*(\delta, \rho)$.

Theorem 1 (Regularization result)**.** *Let the truncation index $m_* = m_*(\delta, \rho)$ be derived as in* (7)*. Then, the inexact Landweber iteration* (14) *truncated at index m_* and updated with the refinement strategy r^δ (satisfying iv) in* Definition 1*) yields for $\alpha(\delta,\rho) = 1/m_*(\delta,\rho)$ a regularization method T_α, which is for all $\nu > 0$ and $0 < \beta < 2/\|\boldsymbol{S}\|^2$ order optimal with constant $\|F\| + c_0$, where c_0 is the constant in* (4) *that can be achieved for classical Landweber regularization.*

Proof. By Lemma 2, it follows that

$$\|\boldsymbol{f}_{m_*}^\delta - \tilde{\boldsymbol{f}}_{m_*}^\delta\| \leq \delta^{\nu/(\nu+1)} \rho^{1/(\nu+1)}.$$

Consequently, by *iv)* in Definition 1 and (4) we obtain

$$\begin{aligned} \|T_\alpha g^\delta - f^\dagger\| &\leq \|F^* \tilde{\boldsymbol{f}}_{m_*}^\delta - F^* \boldsymbol{f}_{m_*}^\delta\| + \|f_{m_*}^\delta - f^\dagger\| \\ &\leq (\|F\| + c_0)\delta^{\nu/(\nu+1)} \rho^{1/(\nu+1)}. \end{aligned}$$

□

4 Regularization theory by a-posteriori parameter rules

As reviewed in Section 2.1, the exact Landweber iteration (6) combined with the discrepancy principle (8)+(9) yields a regularization procedure for problem (1). Within this section, we prove that this result still holds if the approximation is computed by the inexact Landweber iteration (14).

4.1 Adaptive evaluation of the residual discrepancy

The application of the discrepancy principle (8)+(9) requires a frequent evaluation of the residual discrepancy $\|Lf_m^\delta - g^\delta\|$. Therefore, we have to propose a function that is numerical implementable and approximates the residual, preferably in terms of $\boldsymbol{APPLY}$, $\boldsymbol{RHS}$ and $\boldsymbol{COARSE}$.

Definition 2. *For some $g \in \mathcal{H}$, $\boldsymbol{f} \in \ell_2$ and some integer $m \geq 0$ the approximate residual discrepancy $\boldsymbol{RES}$ is defined by*

$$(\boldsymbol{RES}_m[\boldsymbol{f}, g])^2 := \langle \boldsymbol{APPLY}_m[\boldsymbol{f}], \boldsymbol{f}\rangle - 2\langle \boldsymbol{RHS}_m[g], \boldsymbol{f}\rangle + \|g\|^2. \tag{17}$$

The following lemma gives a result on the distance between the exact function space residual discrepancy $\|Lf - g\|$ and its inexact version $\boldsymbol{RES}_m[\boldsymbol{f}, g]$.

Lemma 3. *For $\boldsymbol{f} \in \ell_2$ with $F\boldsymbol{f} = f$, $g \in \mathcal{H}$ and some integer $m \geq 0$ holds*

$$|\; \|Lf - g\|^2 - (\boldsymbol{RES}_m[\boldsymbol{f}, g])^2 \;| \leq (\varepsilon_m^A + 2\varepsilon_m^R)\|\boldsymbol{f}\|. \tag{18}$$

This results follows easily by straightforward computations.

4.2 Decay of approximation errors

In order to achieve convergence of the inexact iteration (14), we follow basic ideas proposed in [25]. Therefore, we have to elaborate under which conditions a monotone decay of the approximation errors $\|\tilde{\boldsymbol{f}}_{m+1}^\delta - \boldsymbol{f}^\dagger\|$ can be ensured.

Lemma 4. *Let $\delta > 0$, $0 < c < 1$, $0 < \beta < 2/(3\|\boldsymbol{S}\|)$ and $m_0 \geq 1$. If there exists for $0 \leq m \leq m_0$ a refinement strategy $r^\delta(m)$ such that the approximate discrepancies $\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}_m^\delta, g^\delta]$ fulfill*

$$c(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}_m^\delta, g^\delta])^2 > \frac{\delta^2 + C_{r^\delta(m)}(\tilde{\boldsymbol{f}}_m^\delta)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}\;, \tag{19}$$

then, for $0 \leq m \leq m_0$, the approximation errors $\|\tilde{\boldsymbol{f}}_m^\delta - \boldsymbol{f}^\dagger\|$ decrease monotonically. The constant $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}_m^\delta)$, which is defined below in (59)*, depends on the iterate $\tilde{\boldsymbol{f}}_m^\delta$, the generalized solution $\boldsymbol{f}^\dagger$, $\|\boldsymbol{S}\|$, β and the error tolerances $\varepsilon_{r^\delta(m)}^{\{C,A,R\}}$.*

The proof of this lemma can be found in Appendix.

4.3 Convergence of the inexact Landweber iterates for exact data

As an observation, Lemma 4 holds in particular for exact data, i.e. for $g^\delta = g$ with $\delta = 0$. In this case, $\tilde{\boldsymbol{f}}_m^\delta$ can be replaced by $\tilde{\boldsymbol{f}}_m$ and r^δ by r and therefore, quantity (59) reduces to $C_{r(m)}(\tilde{\boldsymbol{f}}_m)$. Hence, the condition to achieve a monotone decay of $\|\tilde{\boldsymbol{f}}_m - \boldsymbol{f}^\dagger\|$ simplifies to

$$c(\boldsymbol{RES}_{r(m)}[\tilde{\boldsymbol{f}}_m, g])^2 \geq \frac{C_{r(m)}(\tilde{\boldsymbol{f}}_m)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}\;, \tag{20}$$

which can always be achieved if $C_{r(m)}(\tilde{\boldsymbol{f}}_m)$ is chosen accordingly. In order to prove convergence, we follow the suggested proceeding in [25] and introduce an updating rule (U) for the refinement strategy r:

U(i) Let $r(0)$ be the smallest integer ≥ 0 with

$$c(\boldsymbol{RES}_{r(0)}[\tilde{\boldsymbol{f}}_0, g])^2 \geq \frac{C_{r(0)}(\tilde{\boldsymbol{f}}_0)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{21}$$

if $r(0)$ with (21) does not exist, stop the iteration, set $m_0 = 0$.

U(ii) if for $m \geq 1$

$$c(\boldsymbol{RES}_{r(m-1)}[\tilde{\boldsymbol{f}}_m, g])^2 \geq \frac{C_{r(m-1)}(\tilde{\boldsymbol{f}}_m)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{22}$$

set $r(m) = r(m-1)$

U(iii) if

$$c(\boldsymbol{RES}_{r(m-1)}[\tilde{\boldsymbol{f}}_m, g])^2 < \frac{C_{r(m-1)}(\tilde{\boldsymbol{f}}_m)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{23}$$

set $r(m) = r(m-1) + j$, where j is the smallest integer with

$$c(\boldsymbol{RES}_{r(m-1)+j}[\tilde{\boldsymbol{f}}_m, g])^2 \geq \frac{C_{r(m-1)+j}(\tilde{\boldsymbol{f}}_m)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{24}$$

U(iv) if there is no integer j with (24), then stop the iteration, set $m_0 = m$.

Lemma 5. *Let $\delta = 0$ (i.e. $g^\delta = g$) and $\{\tilde{\boldsymbol{f}}_m\}_{m\in\mathbb{N}}$ be the sequence of iterates derived by* (14). *Assume the updating rule* (U) *for r was applied. Then, if the iteration never stops,*

$$\sum_{m=0}^{\infty} (\boldsymbol{RES}_{r(m)}[\tilde{\boldsymbol{f}}_m, g])^2 \leq \frac{1}{\beta(1-c)\big(1 - \frac{3}{2}\beta\|\boldsymbol{S}\|\big)} \|\tilde{\boldsymbol{f}}_0 - \boldsymbol{f}^\dagger\|^2. \tag{25}$$

If the iteration stops after m_0 steps,

$$\sum_{m=0}^{m_0-1} (\boldsymbol{RES}_{r(m)}[\tilde{\boldsymbol{f}}_m, g])^2 \leq \frac{1}{\beta(1-c)\big(1 - \frac{3}{2}\beta\|\boldsymbol{S}\|\big)} \|\tilde{\boldsymbol{f}}_0 - \boldsymbol{f}^\dagger\|^2. \tag{26}$$

Proof. Suppose first that the iteration never stops. Then, updating rule (U) yields for any $m \geq 0$

$$c(\boldsymbol{RES}_{r(m)}[\tilde{\boldsymbol{f}}_m, g])^2 \geq \frac{C_{r(m)}(\tilde{\boldsymbol{f}}_m)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}. \tag{27}$$

Therefore, it follows from (60) for $\delta = 0$,

$$(\boldsymbol{RES}_{r(m)}[\tilde{\boldsymbol{f}}_m, g])^2 \leq \frac{1}{\beta(1-c)(\frac{3}{2}\beta\|\boldsymbol{S}\| - 1)} \big(\|\tilde{\boldsymbol{f}}_m - \boldsymbol{f}^\dagger\|^2 - \|\tilde{\boldsymbol{f}}_{m+1} - \boldsymbol{f}^\dagger\|^2\big) \tag{28}$$

and summing with respect to m yields (25). If, on the other hand, the iteration terminates after m_0 steps, updating rule (27) holds for $0 \leq m \leq m_0 - 1$. Consequently, one can apply again (28) and summing over the first m_0 summands yields the result. □

Combining the monotone decay of the approximation errors and the uniform boundedness of the accumulated residual discrepancies enables us to prove strong convergence of the inexact iteration (14) towards a solution of (1) for exact data $g^\delta = g$.

Theorem 2. *Let $f^\dagger$ denote a solution of the inverse problem* (1) *and let $\boldsymbol{f}^\dagger$ be the sequence of associated frame coefficients, i.e. $f^\dagger = F^*\boldsymbol{f}^\dagger$. Suppose $\tilde{\boldsymbol{f}}_m$ is computed by the inexact Landweber iteration* (14) *with exact data g in combination with updating rule* (U) *for the refinement strategy r. Then, for $\tilde{\boldsymbol{f}}_0$ arbitrarily chosen, the sequence $\tilde{\boldsymbol{f}}_m$ converges in norm towards $\boldsymbol{f}^\dagger$, i.e.*

$$\lim_{m\to\infty} \tilde{\boldsymbol{f}}_m \to \boldsymbol{f}^\dagger.$$

Of course, the same holds true in the function space topology, i.e. $F^*\tilde{\boldsymbol{f}}_m =: \tilde{f}_m \to f^\dagger$.

Proof. First, we consider the case in which the iteration stops after m_0 steps. Thanks to (59) it follows that $C_i(\tilde{\boldsymbol{f}}_{m_0}) \to 0$ as $i \to \infty$. Therefore, according to the updating rule (U), one has for all $j \geq 1$

$$0 \leq c(\boldsymbol{RES}_{r(m_0-1)+j}[\tilde{\boldsymbol{f}}_{m_0}, g])^2 < \frac{C_{r(m_0-1)+j}(\tilde{\boldsymbol{f}}_{m_0})}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|} \to 0 \text{ as } j \to \infty\ .$$

Moreover, by Lemma 3 formula (18) one has for any $i \geq 0$

$$\|L\tilde{f}_{m_0} - g\|^2 \leq (\boldsymbol{RES}_i[\tilde{\boldsymbol{f}}_{m_0}, g])^2 + (\varepsilon_i^A + 2\varepsilon_i^R)\|\tilde{\boldsymbol{f}}_{m_0}\| \to 0 \text{ as } i \to \infty\ ,$$

and therefore $g = LF^*\tilde{\boldsymbol{f}}_{m_0}$. Consider now the case in which the iteration never stops. For norm convergence it is sufficient to show that

$$e_k := \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_k$$

is a Cauchy sequence. From the updating rule (U) and Lemma 4 it is clear that $\|e_k\|$ forms a decreasing sequence. For $n \geq k$ select an index $l = l(k,n)$ such that $k \leq l \leq n$ and that

$$\boldsymbol{RES}_{r(l)}[\tilde{\boldsymbol{f}}_l, g] := \min_{k\leq i\leq n}\{\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g]\} \leq \boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g] \text{ for } k \leq i \leq n.$$

Since

$$\|e_n - e_k\| \leq \|e_n - e_l\| + \|e_l - e_k\|\ ,$$

$$\begin{aligned}\|e_n - e_l\|^2 &= \|e_n\|^2 - \|e_l\|^2 + 2\langle e_l - e_n, e_l\rangle \\ \|e_l - e_k\|^2 &= \|e_k\|^2 - \|e_l\|^2 + 2\langle e_l - e_k, e_l\rangle\end{aligned}$$

and since $\|e_i\|$ are monotonically decreasing, i.e. there is some $\varepsilon \geq 0$ with $\|e_i\| \to \varepsilon$ as $i \to \infty$, it is sufficient to verify

$$|\langle e_l - e_n, e_l\rangle| \to 0 \text{ and } |\langle e_l - e_k, e_l\rangle| \to 0 \text{ for } k \to \infty.$$

To this end, consider

$$\begin{aligned}
\frac{1}{\beta}|\langle e_l - e_k, e_l\rangle| &= \frac{1}{\beta}|\langle \tilde{\boldsymbol{f}}_k - \tilde{\boldsymbol{f}}_l, \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_l\rangle| \le \frac{1}{\beta}\sum_{i=k}^{l-1}|\langle \tilde{\boldsymbol{f}}_{i+1} - \tilde{\boldsymbol{f}}_i, \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_l\rangle| \\
&\le \sum_{i=k}^{l-1}|\langle \boldsymbol{RHS}_{r(i)}[\boldsymbol{g}] - \boldsymbol{APPLY}_{r(i)}[\boldsymbol{S}, \tilde{\boldsymbol{f}}_i], \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_l\rangle| \\
&\le \sum_{i=k}^{l-1}\big[|\langle \boldsymbol{RHS}_{r(i)}[\boldsymbol{g}] - \boldsymbol{g}, \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_l\rangle| \\
&\qquad + |\langle \boldsymbol{S}\tilde{\boldsymbol{f}}_i - \boldsymbol{APPLY}_{r(i)}[\boldsymbol{S}, \tilde{\boldsymbol{f}}_i], \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_l\rangle| \\
&\qquad + |\langle \boldsymbol{g} - \boldsymbol{S}\tilde{\boldsymbol{f}}_i, \boldsymbol{f}^\dagger - \boldsymbol{f}_l\rangle|\big] \\
&\le \sum_{i=k}^{l-1}\big[(\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})\|\boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_l\| + \|g - LF^*\tilde{\boldsymbol{f}}_i\|\|g - LF^*\tilde{\boldsymbol{f}}_l\|\big] \\
&\le \sum_{i=k}^{l-1}\big[(\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})\|\boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_i\| \\
&\qquad + \frac{1}{2}(\|g - LF^*\tilde{\boldsymbol{f}}_i\|^2 + \|g - LF^*\tilde{\boldsymbol{f}}_l\|^2)\big] \\
&\le \sum_{i=k}^{l-1}\big[(\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})\|\boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_i\| \\
&\qquad + \frac{1}{2}((\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g])^2 + (2\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})\|\tilde{\boldsymbol{f}}_i\| \\
&\qquad + (\boldsymbol{RES}_{r(l)}[\tilde{\boldsymbol{f}}_l, g])^2 + (2\varepsilon^R_{r(l)} + \varepsilon^A_{r(l)})\|\tilde{\boldsymbol{f}}_l\|)\big] \quad .
\end{aligned} \tag{29}$$

¿From (59) one has for all $k \le i \le l-1$,

$$(\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})\|\boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_i\| \le (\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})(\|\boldsymbol{f}^\dagger\| + \|\tilde{\boldsymbol{f}}_i\|) \le C_{r(i)}(\tilde{\boldsymbol{f}}_i)$$

and

$$(2\varepsilon^R_{r(i)} + \varepsilon^A_{r(i)})\|\tilde{\boldsymbol{f}}_i\| \le C_{r(i)}(\tilde{\boldsymbol{f}}_i) \ .$$

Moreover, updating rule (U) yields

$$C_{r(i)}(\tilde{\boldsymbol{f}}_i) \le c(1 - \frac{3}{2}\beta\|\boldsymbol{S}\|)(\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g])^2.$$

Consequently, since $\boldsymbol{RES}_{r(l)}[\tilde{\boldsymbol{f}}_l, g] \le \boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g]$, estimate (29) can be rewritten,

$$\begin{aligned}
\frac{1}{\beta}|\langle e_l - e_k, e_l\rangle| &\le \sum_{i=k}^{l-1}\Bigg(C_{r(i)}(\tilde{\boldsymbol{f}}_i) + \frac{1}{2}\Bigg[(\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g])^2 + C_{r(i)}(\tilde{\boldsymbol{f}}_i) \\
&\qquad + (\boldsymbol{RES}_{r(l)}[\tilde{\boldsymbol{f}}_l, g])^2 + C_{r(l)}(\tilde{\boldsymbol{f}}_l)\Bigg]\Bigg) \\
&\le \sum_{i=k}^{l-1}\Bigg((\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g])^2 + 2c(1 - \frac{3}{2}\beta\|\boldsymbol{S}\|)(\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g])^2\Bigg) \\
&= const\sum_{i=k}^{l-1}(\boldsymbol{RES}_{r(i)}[\tilde{\boldsymbol{f}}_i, g])^2 \ ,
\end{aligned}$$

which tends to zero as $k \to \infty$ thanks to Lemma 5. Analogously it can be shown that

$$|\langle e_l - e_j, e_l \rangle| \to 0 \quad \text{as} \quad k \to \infty.$$

Thus, the sequence $e_k = \boldsymbol{f}^\dagger - \tilde{\boldsymbol{f}}_k$ forms a Cauchy sequence with $\|e_k\| \to 0$ and therefore,

$$\tilde{\boldsymbol{f}}_k \to \boldsymbol{f}^\dagger \quad \text{as} \quad k \to \infty$$

implying

$$\|\tilde{f}_k - f^\dagger\| \leq \|F^*\| \|\tilde{\boldsymbol{f}}_k - \boldsymbol{f}^\dagger\| \to 0 \quad \text{as} \quad k \to \infty \; .$$

□

4.4 Convergence of the inexact Landweber iterates for noisy data

The convergence proof of the truncated inexact Landweber method for noisy data essentially relies on a comparison between noise free and noisy iterations. To apply this comparison principle, one has to analyze the δ-dependence of $\boldsymbol{COARSE}$, $\boldsymbol{APPLY}$ and $\boldsymbol{RHS}$ (in particular for $\delta \to 0$). For our situation, it would be desirable that for a given error level $\|\boldsymbol{v}^\delta - \boldsymbol{v}\| \leq \delta$ the routines (exemplarily stated for $\boldsymbol{COARSE}$) fulfill for any fixed $\varepsilon > 0$

$$\|\boldsymbol{COARSE}_\varepsilon(\boldsymbol{v}^\delta) - \boldsymbol{COARSE}_\varepsilon(\boldsymbol{v})\| \to 0 \quad \text{as} \quad \delta \to 0. \tag{30}$$

Such a property must hold for all the three routines. As we shall see below, this requires a slight δ-adaptation of the routines that were proposed in [29]; which is here only explicitly demonstrated for $\boldsymbol{COARSE}$ (but can be done analogously for $\boldsymbol{APPLY}$ and $\boldsymbol{RHS}$).

Since the input of $\boldsymbol{COARSE}$ are finite length vectors, we limit the analysis to vectors $\boldsymbol{v}, \boldsymbol{v}^\delta$ that have finitely many non-zero elements. The definition of $\boldsymbol{COARSE}$ as proposed in [29] (with a slight modified ordering of the output entries) is the following

$\boldsymbol{COARSE}_\varepsilon[\boldsymbol{v}] \to \boldsymbol{v}_\varepsilon$

i) Let $\boldsymbol{V}$ be the set of non-zero coefficients of $\boldsymbol{v}$, ordered by their original indexing in $\boldsymbol{v}$. Define $q := \left\lceil \log\left(\frac{(\#\boldsymbol{V})^{1/2}\|\boldsymbol{v}\|}{\varepsilon} \right) \right\rceil$.

ii) Divide the elements of $\boldsymbol{V}$ into bins $\boldsymbol{V}_0, \ldots, \boldsymbol{V}_q$, where for $0 \leq k < q$

$$\boldsymbol{V}_k := \{v \in \boldsymbol{V} : 2^{-k-1}\|\boldsymbol{v}\| < |v| \leq 2^{-k}\|\boldsymbol{v}\|\}, \tag{31}$$

and possible remaining elements are put into $\boldsymbol{V}_q$. Let the elements of a single $\boldsymbol{V}_k$ be also ordered by their original indexing in $\boldsymbol{v}$. Denote the vector obtained by subsequently extracting the elements of $\boldsymbol{V}_0, \ldots, \boldsymbol{V}_q$ by $\gamma(\boldsymbol{v})$.

iii) Create $\boldsymbol{v}_\varepsilon$ by extracting elements from $\gamma(\boldsymbol{v})$ and putting them at the original indices, until the smallest l is found with

$$\|\boldsymbol{v} - \boldsymbol{v}_\varepsilon\|^2 = \sum_{i>l} |\gamma_i(\boldsymbol{v})|^2 < \varepsilon^2. \tag{32}$$

Remark 1. *The integer q in* i) *is chosen such that $\sum_{v \in \boldsymbol{V}_q} |v|^2 < \varepsilon^2$, i.e. the elements of $\boldsymbol{V}_q$ are not used to build $\boldsymbol{v}_\varepsilon$ in* iii)*.*

Remark 2. *Keeping the original order of the coefficients of $\boldsymbol{v}$ in $\boldsymbol{V}_k$, the output vector of $\boldsymbol{COARSE}$ becomes unique. This "natural" ordering does not cause an extra computational cost.*

Remark 3. *Since we aim to show that the noise dependent output vector converges to the noise free output vector for* $\delta \to 0$*, the uniqueness of* $\boldsymbol{COARSE}$ *plays an important role. The main problem is the non-uniqueness of bin sorting processed by* $\boldsymbol{COARSE}$ *(due to the noise). This, unfortunately, leads to the problem that the index in* $\gamma(\boldsymbol{v}^\delta)$ *of some noisy element* v_i^δ *can differ to the index in* $\gamma(\boldsymbol{v})$ *of it's noise free version* v_i*. To overcome this drawback, at least for sufficiently small* δ*, we define a noise dependent version* $\boldsymbol{COARSE}^\delta$*.*

$\boldsymbol{COARSE}_\varepsilon^\delta[\boldsymbol{v}^\delta] \to \boldsymbol{v}_\varepsilon^\delta$

i) Let $\boldsymbol{V}^\delta$ be the set of non-zero coefficients of $\boldsymbol{v}^\delta$ ordered by their indexing in $\boldsymbol{v}^\delta$. Define $q^\delta := \left\lceil \log\left(\frac{(\#\boldsymbol{V}^\delta)^{1/2}(\|\boldsymbol{v}^\delta\|+\delta)}{\varepsilon}\right)\right\rceil$.

ii) Divide the elements of $\boldsymbol{V}^\delta$ into bins $\boldsymbol{V}_0^\delta, \dots, \boldsymbol{V}_{q^\delta}^\delta$, where for $0 \le k < q^\delta$

$$\boldsymbol{V}_k^\delta := \{v_i^\delta \in \boldsymbol{V}^\delta : 2^{-k-1}(\|\boldsymbol{v}_i^\delta\| + \delta) + \delta < |v_i^\delta| \le 2^{-k}(\|\boldsymbol{v}^\delta\| + \delta) + \delta\}, \tag{33}$$

and possible remaining elements are put into $\boldsymbol{V}_{q^\delta}^\delta$. Again, let the elements of a single $\boldsymbol{V}_k^\delta$ be ordered by their indexing in $\boldsymbol{v}^\delta$. Denote the vector obtained by the bin sorting process by $\gamma^\delta(\boldsymbol{v}^\delta)$.

iii) Create $\boldsymbol{v}_\varepsilon^\delta$ by extracting elements from $\gamma^\delta(\boldsymbol{v}^\delta)$ and putting them on the original places, until the first index l^δ is found with

$$\|\boldsymbol{v}^\delta - \boldsymbol{v}_\varepsilon^\delta\|^2 = \|\boldsymbol{v}^\delta\|^2 - \sum_{1\le i\le l^\delta} |\gamma_i^\delta(\boldsymbol{v}^\delta)|^2 < \varepsilon^2 - (l^\delta + 1)\delta(2\|\boldsymbol{v}^\delta\| + \delta). \tag{34}$$

The latter definition of $\boldsymbol{COARSE}^\delta$ enables us to achieve property (30).

Lemma 6. *Given* $\varepsilon > 0$ *and* $\delta > 0$*. For arbitrary finite length vectors* $\boldsymbol{v}$*,* $\boldsymbol{v}^\delta \in \ell_2$ *with* $\|\boldsymbol{v}^\delta - \boldsymbol{v}\| \le \delta$*, the routine* $\boldsymbol{COARSE}^\delta$ *is convergent in the sense that*

$$\|\boldsymbol{COARSE}_\varepsilon^\delta[\boldsymbol{v}^\delta] - \boldsymbol{COARSE}_\varepsilon[\boldsymbol{v}]\| \to 0 \quad \text{as} \quad \delta \to 0. \tag{35}$$

Proof. For $\delta = 0$, it follows from the definition that $\boldsymbol{COARSE}^\delta$ coincides with $\boldsymbol{COARSE}$.

Let $\delta > 0$. For sufficiently small δ, the number $\#\boldsymbol{V}^\delta$ of the coefficients of $\boldsymbol{v}^\delta$ is greater or equal than $\#\boldsymbol{V}$. In fact, for all δ satisfying $\delta \le \delta_0 < \inf_{v_i \in \boldsymbol{V}} |v_i|$ we have due to $\|v_i^\delta - v_i\| \le \delta$ the relation $|v_i^\delta| \ge |v_i| - \delta > 0$. Therefore, due to the definition of q^δ, it follows

$$q^\delta = \left\lceil \log\left(\frac{(\#\boldsymbol{V}^\delta)^{1/2}(\|\boldsymbol{v}^\delta\| + \delta)}{\varepsilon}\right)\right\rceil \ge \left\lceil \log\left(\frac{(\#\boldsymbol{V})^{1/2}\|\boldsymbol{v}\|}{\varepsilon}\right)\right\rceil = q.$$

In what follows we show that for $\delta \le \delta_0$ the elements of any $\boldsymbol{V}_k^\delta$ (for $0 \le k \le q-1$) are exactly the noisy counterparts of the elements of $\boldsymbol{V}_k$ (due to the indexing assumption they are ordered in the same way). To this end, suppose that v_i^δ belongs for some $0 \le k \le q-1 < q^\delta$ to $\boldsymbol{V}_k^\delta$. Then, with the left inequality of (33) and the relations $||v_i| - |v_i^\delta|| \le \delta$ and $|\|\boldsymbol{v}\| - \|\boldsymbol{v}^\delta\|| \le \delta$, we have

$$|v_i| \ge |v_i^\delta| - \delta > 2^{-(k+1)}(\|\boldsymbol{v}^\delta\| + \delta) + \delta - \delta \ge 2^{-(k+1)}\|\boldsymbol{v}\|.$$

By the right inequality of (33) we achieve for $\delta \to 0$ that $|v_i| \le 2^{-k}\|\boldsymbol{v}\|$. Consequently, by (31), we obtain $v_i \in \boldsymbol{V}_k$. Conversely, if $v_i \in \boldsymbol{V}_k$, the counterpart v_i^δ must be in $\boldsymbol{V}_k^\delta$, otherwise it would contradict to the recently proven result.

Next, we prove that for δ small enough the thresholding index l in $\boldsymbol{COARSE}$ coincides with l^δ in $\boldsymbol{COARSE}^\delta$. Let $\boldsymbol{v}_\varepsilon$ be the output of $\boldsymbol{COARSE}$. Then, l is the smallest integer with

$$\|\boldsymbol{v} - \boldsymbol{v}_\varepsilon\|^2 = \|\boldsymbol{v}\|^2 - \sum_{1\le i\le l} |\gamma_i(\boldsymbol{v})|^2 < \varepsilon^2. \tag{36}$$

Since $\|\boldsymbol{v}^\delta\|$ is bounded for $\delta \to 0$ and l is fixed for a fixed $\boldsymbol{v}$, there is some $\delta_1 \le \delta_0$ such that for all $\delta \le \delta_1$

$$\sum_{1\le i\le l} |\gamma_i(\boldsymbol{v})|^2 > \|\boldsymbol{v}\|^2 - \varepsilon^2 + 2(l+1)\delta(2\|\boldsymbol{v}^\delta\| + \delta). \tag{37}$$

The l-th element of $\gamma(\boldsymbol{v})$ is in one of the containers $\boldsymbol{V}_k$ with $0 \le k \le q-1$ (see Remark 1), the first l elements of $\gamma^\delta(\boldsymbol{v}^\delta)$ are exactly the noisy equivalents of the first l values of $\gamma(\boldsymbol{v})$, i.e. $|\gamma_i^\delta(\boldsymbol{v}^\delta) - \gamma_i(\boldsymbol{v})| \le \delta$, $1 \le i \le l$. Consequently, we get for $1 \le i \le l$

$$\Big| |\gamma_i^\delta(\boldsymbol{v}^\delta)|^2 - |\gamma_i(\boldsymbol{v})|^2 \Big| \le \delta(|\gamma_i^\delta(\boldsymbol{v}^\delta)| + |\gamma_i(\boldsymbol{v})|) \le \delta(\|\boldsymbol{v}^\delta\| + \|\boldsymbol{v}\|) \le \delta(2\|\boldsymbol{v}^\delta\| + \delta). \tag{38}$$

Analogously,

$$\Big| \|\boldsymbol{v}^\delta\|^2 - \|\boldsymbol{v}\|^2 \Big| = \Big| \|\boldsymbol{v}^\delta\| - \|\boldsymbol{v}\| \Big| (\|\boldsymbol{v}^\delta\| + \|\boldsymbol{v}\|) \le \delta(\|\boldsymbol{v}^\delta\| + \|\boldsymbol{v}\|) \le \delta(2\|\boldsymbol{v}^\delta\| + \delta). \tag{39}$$

Combining estimates (37)-(39), we have

$$\begin{aligned}
\sum_{1\le i\le l} |\gamma_i^\delta(\boldsymbol{v}^\delta)|^2 &\overset{(38)}{\ge} \sum_{1\le i\le l} |\gamma_i(\boldsymbol{v})|^2 - l\delta(2\|\boldsymbol{v}^\delta\| + \delta) \\
&\overset{(37)}{>} \|\boldsymbol{v}\|^2 - \varepsilon^2 + 2(l+1)\delta(2\|\boldsymbol{v}^\delta\| + \delta) - l\delta(2\|\boldsymbol{v}^\delta\| + \delta) \\
&\overset{(39)}{\ge} \|\boldsymbol{v}^\delta\|^2 - \delta(2\|\boldsymbol{v}^\delta\| + \delta) - \varepsilon^2 + (l+2)\delta(2\|\boldsymbol{v}^\delta\| + \delta) \\
&= \|\boldsymbol{v}^\delta\|^2 - \varepsilon^2 + (l+1)\delta(2\|\boldsymbol{v}^\delta\| + \delta).
\end{aligned} \tag{40}$$

Therefore (by the last estimate (40)) l satisfies (34), i.e. $l^\delta \le l$. Conversely supposing that $l^\delta < l$ yields

$$\begin{aligned}
\sum_{1\le i\le l^\delta} |\gamma_i(\boldsymbol{v})|^2 &\overset{(38)}{\ge} \sum_{1\le i\le l^\delta} |\gamma_i(\boldsymbol{v}^\delta)|^2 - l^\delta\delta(2\|\boldsymbol{v}^\delta\| + \delta) \\
&\overset{(34)}{>} \|\boldsymbol{v}^\delta\|^2 - \varepsilon^2 + (l^\delta+1)\delta(2\|\boldsymbol{v}^\delta\| + \delta) - l^\delta\delta(2\|\boldsymbol{v}^\delta\| + \delta) \\
&\overset{(39)}{\ge} \|\boldsymbol{v}\|^2 - \delta(2\|\boldsymbol{v}^\delta\| + \delta) - \varepsilon^2 + \delta(2\|\boldsymbol{v}^\delta\| + \delta) \\
&= \|\boldsymbol{v}\|^2 - \varepsilon^2,
\end{aligned}$$

which contradicts to the definition of l in $\boldsymbol{COARSE}$. Consequently, we summarize that for sufficiently small $\delta > 0$ the labels l^δ and l coincide; hence

$$\|\boldsymbol{v}_\varepsilon^\delta - \boldsymbol{v}_\varepsilon\| \le \|\boldsymbol{v}^\delta - \boldsymbol{v}\| \le \delta,$$

i.e. $\boldsymbol{COARSE}^\delta$ converges in the sense of (35). □

In order to prove convergence of the inexact iteration, we introduce (inspired by the proceeding in [25]) as in the noise free case an updating rule which we denote (D). This updating rule is based on the refinement strategy $r^\delta(m)$.

D(i) Let $r^\delta(0)$ be the smallest integer ≥ 0 with

$$c(\boldsymbol{RES}_{r^\delta(0)}[\tilde{\boldsymbol{f}}_0^\delta, g^\delta])^2 \geq \frac{\delta^2 + C_{r^\delta(0)}(\tilde{\boldsymbol{f}}_0^\delta)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{41}$$

if $r^\delta(0)$ does not exist, stop the iteration, set $m_* = 0$.

D(ii) if for $m \geq 1$

$$c(\boldsymbol{RES}_{r^\delta(m-1)}[\tilde{\boldsymbol{f}}_m^\delta, g^\delta])^2 \geq \frac{\delta^2 + C_{r^\delta(m-1)}(\tilde{\boldsymbol{f}}_m^\delta)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{42}$$

set $r^\delta(m) = r^\delta(m-1)$

D(iii) if

$$c(\boldsymbol{RES}_{r^\delta(m-1)}[\tilde{\boldsymbol{f}}_m^\delta, g^\delta])^2 < \frac{\delta^2 + C_{r^\delta(m-1)}(\tilde{\boldsymbol{f}}_m^\delta)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|}, \tag{43}$$

set $r^\delta(m) = r^\delta(m-1) + j$, where j is the smallest integer with

$$c(\boldsymbol{RES}_{r^\delta(m-1)+j}[\tilde{\boldsymbol{f}}_m^\delta, g^\delta])^2 \geq \frac{\delta^2 + C_{r^\delta(m-1)+j}(\tilde{\boldsymbol{f}}_m^\delta)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|} \tag{44}$$

and

$$C_{r^\delta(m-1)+j}(\tilde{\boldsymbol{f}}_m^\delta) > c_1\delta^2. \tag{45}$$

D(iv) if (43) holds and no j with (44),(45) exists, then stop the iteration, set $m_*^\delta = m$.

Theorem 3. *Let $f^\dagger$ be again a solution of* (1) *for exact data $g \in \operatorname{Ran} L$. Suppose that for any $\delta > 0$ and g^δ with $\|g^\delta - g\| \leq \delta$, the adaptive approximation $\tilde{\boldsymbol{f}}_m^\delta$ is derived by the inexact Landweber iteration* (14) *in combination with updating rule* (D) *for the refinement strategy r^δ and the stopping index $m_* = m_*^\delta$. Then, the family of T_α defined as*

$$T_\alpha g^\delta := F^* \tilde{\boldsymbol{f}}_{m_*^\delta}^\delta \quad \text{with} \quad \alpha = \alpha(\delta, g^\delta) = \frac{1}{m_*^\delta}$$

yields a regularization of the ill-posed operator L, i.e.

$$\|T_\alpha g^\delta - f^\dagger\|_X = \|\tilde{f}_{m_*^\delta}^\delta - f^\dagger\|_X \to 0 \quad \text{as} \quad \delta \to 0, \tag{46}$$

with $\tilde{f}_{m_*^\delta}^\delta := F^* \tilde{\boldsymbol{f}}_{m_*^\delta}^\delta$.

Proof. The proof goes as follows (similar to the idea proposed in [25]): Exploiting induction with respect to the iteration index m, we first show that for sufficiently small δ the refinement strategy $r^\delta(m)$ in the noisy case is equal to the refinement strategy $r(m)$ in the noise-free case. Therewith, for some fixed m, $\tilde{\boldsymbol{f}}_m^\delta \to \tilde{\boldsymbol{f}}_m$ as $\delta \to 0$. Second, we consider convergence properties of the truncation indices $m_*^{\delta_n}$ for some sequence δ_n tending to 0 as $n \to \infty$. As a result of this analysis we are able to deduce that in all possible cases the corresponding truncated adaptive iterations $\tilde{\boldsymbol{f}}_{m_*^{\delta_n}}^{\delta_n}$ converge with $n \to \infty$.

- We compare the noise free with the noisy iteration, i.e. $\tilde{\boldsymbol{f}}_{m+1}$ defined through (14)+(U) and $\tilde{\boldsymbol{f}}^{\delta}_{m+1}$ defined through (14)+(D). We assume the initial values of both iterations to be the same, i.e. $\tilde{\boldsymbol{f}}^{\delta}_0 = \tilde{\boldsymbol{f}}_0$. For sufficiently small values of δ we first aim to show $r^{\delta}(0) = r(0)$. Suppose for some $m \geq 0$ that $\tilde{\boldsymbol{f}}^{\delta}_m \to \tilde{\boldsymbol{f}}_m$ as $\delta \to 0$. Then, $C_{r^{\delta}(m)}(\tilde{\boldsymbol{f}}^{\delta}_m) \to C_{r^{\delta}(m)}(\tilde{\boldsymbol{f}}_m)$ as well as $\boldsymbol{RES}_{r^{\delta}(m)}[\tilde{\boldsymbol{f}}^{\delta}_m, g^{\delta}] \to \boldsymbol{RES}_{r^{\delta}(m)}[\tilde{\boldsymbol{f}}_m, g]$ (which holds true due to convergence properties of the modified routines $\boldsymbol{APPLY}$ and $\boldsymbol{RHS}$). In particular, due to the initial value condition $\tilde{\boldsymbol{f}}^{\delta}_0 = \tilde{\boldsymbol{f}}_0$ we have $C_{r^{\delta}(0)}(\tilde{\boldsymbol{f}}^{\delta}_0) \to C_{r^{\delta}(0)}(\tilde{\boldsymbol{f}}_0)$ and $\boldsymbol{RES}_{r^{\delta}(0)}[\tilde{\boldsymbol{f}}^{\delta}_0, g^{\delta}] \to \boldsymbol{RES}_{r(0)}[\tilde{\boldsymbol{f}}_0, g]$. With the help of (D) we deduce for $\delta \to 0$ that $r^{\delta}(0)$ is the smallest integer ≥ 0 with

$$c(\boldsymbol{RES}_{r^{\delta}(0)}[\tilde{\boldsymbol{f}}_0, g])^2 \geq \frac{C_{r^{\delta}(0)}(\tilde{\boldsymbol{f}}_0)}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|},$$

which is equivalent to the definition of $r(0)$ from (U). Consequently, for sufficiently small δ this means $r^{\delta}(0) = r(0)$ and thus convergence of the first iterates $\tilde{\boldsymbol{f}}^{\delta}_1 \to \tilde{\boldsymbol{f}}_1$ for $\delta \to 0$. Assume now for some $m \geq 1$ that $r^{\delta}(m-1) = r(m-1)$ and $\tilde{\boldsymbol{f}}^{\delta}_m \to \tilde{\boldsymbol{f}}_m$ as $\delta \to 0$. The same argument as for $m = 0$ applies and we conclude that $r^{\delta}(m) = r(m)$ for sufficiently small δ, and therefore $\tilde{\boldsymbol{f}}^{\delta}_{m+1} \to \tilde{\boldsymbol{f}}_{m+1}$.

- Second, we prove convergence in sequence space, i.e.

$$\|\tilde{\boldsymbol{f}}^{\delta}_{m^{\delta}_*} - \boldsymbol{f}^{\dagger}\| \to 0 \text{ as } \delta \to 0 \ . \tag{47}$$

  To this end, let δ_n be a positive sequence with $\delta_n \to 0$ as $n \to \infty$.

  a) Consider first the case in which $m^{\delta_n}_*$ converges to some integer $\tilde{m}$. Then, for sufficiently large n it holds $m^{\delta_n}_* = \tilde{m}$. We denote $m_n := m^{\delta_n}_*$. Since for a fixed k, $m_n - 1$ is the greatest integer with

$$C_{r^{\delta}(m_n-1)}(\tilde{\boldsymbol{f}}^{\delta}_{m_n-1}) > c_1\delta_n \text{ and } c(\boldsymbol{RES}_{r(m_n-1)}[\tilde{\boldsymbol{f}}^{\delta}_{m_n-1}, g^{\delta}])^2 \geq \frac{\delta_n^2 + C_{r(m_n-1)}(\tilde{\boldsymbol{f}}^{\delta}_{m_n-1})}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|},$$

  we obtain with $n \to \infty$ that $\tilde{m} - 1$ is the greatest integer with

$$c\boldsymbol{RES}_{r(\tilde{m}-1)}[\tilde{\boldsymbol{f}}_{\tilde{m}-1}, g]^2 \geq \frac{C_{r(\tilde{m}-1)}(\tilde{\boldsymbol{f}}_{\tilde{m}-1})}{1 - \frac{3}{2}\beta\|\boldsymbol{S}\|},$$

  which is the definition of the stopping index m_* in the noise-free case. Therefore, for sufficiently large n we have $m^{\delta_n}_* = m_*$ and $\tilde{\boldsymbol{f}}^{\delta_n}_{m^{\delta_n}_*} = \tilde{\boldsymbol{f}}^{\delta_n}_{m_*} \to \tilde{\boldsymbol{f}}_{m_*} = \boldsymbol{f}^{\dagger}$ for $n \to \infty$.

  b) Consider now the case in which $m^{\delta_n}_* \to \infty$ (monotonically) as $n \to \infty$. For $n > l$ we use that the errors $\|\tilde{\boldsymbol{f}}^{\delta}_m - \boldsymbol{f}^{\dagger}\|$ decrease with increasing m. The triangle inequality yields

$$\|\tilde{\boldsymbol{f}}^{\delta_n}_{m_n} - \boldsymbol{f}^{\dagger}\| \leq \|\tilde{\boldsymbol{f}}^{\delta_n}_{m_l} - \boldsymbol{f}^{\dagger}\| \leq \|\tilde{\boldsymbol{f}}^{\delta_n}_{m_l} - \tilde{\boldsymbol{f}}_{m_l}\| + \|\tilde{\boldsymbol{f}}_{m_l} - \boldsymbol{f}^{\dagger}\|.$$

  Since the noise free iterates $\tilde{\boldsymbol{f}}_m$ converge to $\boldsymbol{f}^{\dagger}$, choose l such that $\|\tilde{\boldsymbol{f}}_{m_l} - \boldsymbol{f}^{\dagger}\| \leq \frac{\varepsilon}{2}$. Moreover, since $\tilde{\boldsymbol{f}}^{\delta}_m \to \tilde{\boldsymbol{f}}_m$, choose n large enough such that $\|\tilde{\boldsymbol{f}}^{\delta_n}_{m_l} - \tilde{\boldsymbol{f}}_{m_l}\| \leq \frac{\varepsilon}{2}$. This finally yields

$$\|\tilde{\boldsymbol{f}}^{\delta_n}_{m_n} - \boldsymbol{f}^{\dagger}\| \to 0 \text{ as } n \to \infty. \tag{48}$$

  c) Assume that $\tilde{\boldsymbol{f}}^{\delta_n}_{m_n}$ does not converge to $\boldsymbol{f}^{\dagger}$. Then for any $\varepsilon > 0$ there exists a subsequence $\tilde{\boldsymbol{f}}^{\delta_{n(k)}}_{m_{n(k)}}$ of $\tilde{\boldsymbol{f}}^{\delta_{m_n}}_{m_n}$ with

$$\|\tilde{\boldsymbol{f}}^{\delta_{n(k)}}_{m_{n(k)}} - \boldsymbol{f}^{\dagger}\| > \varepsilon. \tag{49}$$

If $m_{n(k)}$ is bounded, then it contains a convergent subsequence. Proceeding as in a), we end up with a contradiction to (49). If $m_{n(k)}$ is unbounded, then it contains a monotonically increasing unbounded subsequence. Proceeding as in b), we again end up with a contradiction to (49). Consequently, (48) holds, and since the sequence δ_n was chosen arbitrarily, we conclude (47). Due to

$$\|\tilde{f}^{\delta}_{m^{\delta}_*} - f^{\dagger}\|_X = \|F^*(\tilde{\boldsymbol{f}}^{\delta}_{m^{\delta}_*} - \boldsymbol{f}^{\dagger})\|_X \leq \|F^*\|\|\tilde{\boldsymbol{f}}^{\delta}_{m^{\delta}_*} - \boldsymbol{f}^{\dagger}\|_{\ell_2} ,$$

we achieve with (47) convergence in X:

$$\|\tilde{f}^{\delta}_{m^{\delta}_*} - f^{\dagger}\|_X \to 0 \text{ as } \delta \to 0 .$$

□

5 Numerical experiment

Within this section, we give a first experiment for the proposed adaptive regularization scheme which we aim to apply to the inversion of the Radon transform.

5.1 Radon transform

Let $\Omega := [0,1]^2$ and $f \in L_2(\mathbb{R}^2)$ having $\mathrm{supp}(f) \subset \Omega$ the signal to be reconstructed. The linear Radon transform $R : L_2(\mathbb{R}^2) \to L_2(Z)$ is then defined by

$$(Rf)(s,\omega) = \int_{\mathbb{R}^2} f(x)\delta(s - \langle x, \omega\rangle)dx,$$

where $Z := \mathbb{R} \times \boldsymbol{S}$ with $\boldsymbol{S} = \{\omega \in \mathbb{R}^2 : \omega = (\cos\theta, \sin\theta), \theta \in [0,\pi)\}$. For some $g \in L_2(Z)$, the adjoint of R is given by

$$(R^*g)(x) = \int_0^{\pi} g(\langle x, \omega\rangle, \omega)d\theta.$$

The inverse problem is then to find an approximation to the solution of the linear equation

$$Rf = g, \tag{50}$$

for which only noisy data $g^{\delta} \in L_2(T)$ are given with $\|g^{\delta} - g\| < \delta$ ($\delta > 0$ is known). As a fact, the Radon operator R is compact and is therefore not continuously invertible. Consequently, problem (50) is ill-posed and needs to be stabilized/regularized. For the case $g^{\delta} \notin \mathrm{Ran}\, R$, we consider instead of (50) the Gaussian data misfit term

$$\|Rf - g^{\delta}\|$$

for which a minimum has to fulfill the normal equation

$$R^*Rf = R^*g^{\delta}. \tag{51}$$

The operator R^*R is not boundedly invertible and therefore the adaptive approach (as suggested in [29]) for well-posed and symmetric problems is not applicable. For this problem our proposed approach that combines regularization and adaptivity is adequate.

5.2 Sequence space formulation and adaptive approximation

As mentioned above, we emphasize on a wavelet-based discretization of (51). The advantages/capabilities of wavelets in the context of adaptive approximation have been extensively analyzed, e.g. in [7, 9, 30]. As one important result, for many operator equations wavelet-based discretizations yield compressible (or even sparse) stiffness matrices (which are in principle dense when discretizing with respect to a single scale basis). Compressibility typically leads to low cost matrix vector multiplications with minor loss of accuracy and essentially depending on the properties of the operator and the involved wavelet systems. A well studied implementation of such low cost matrix vector multiplication is the routine ***APPLY*** leading (in combination with ***COARSE*** and ***RHS***) to optimal computational complexity. Another advantage of adaptive approximation (as mentioned in Section 2) is the improved convergence rate. It can be achieved only in cases in which the solution has a Besov regularity that is higher than the Sobolev regularity, which is the case for the tomographic reconstruction problem, see Section 2.

In order to study the usefulness (in the sense of optimal computational complexity) of adaptive approximation for solving (51), we recapitulate the basic framework (as elaborated in [7], [9] or [30]). Assume we are given a dual pair of wavelet frames $\Psi, \tilde{\Psi}$ with regularity bounds γ, $\tilde{\gamma}$ and vanishing moments d, $\tilde{d}$. Suppose, moreover, the operator under consideration is bounded and of the order $2t$, i.e. for some $\sigma > 0$ the operator maps continuously between Sobolev spaces $H^{t+\sigma}$ and $H^{-t+\sigma}$. Then, ***APPLY*** performs with optimal complexity if $s^* = \min\{\sigma, \gamma - t, t + \tilde{d}\}/n - 1/2 > 0$, where n denotes the spatial dimension. For a detailed analysis we refer e.g. to [7], [9].

Let us now construct a setup in which the proposed adaptive approximation of a solution of (51) meets the latter requirements. At first, we have to introduce adequate Sobolev spaces that can be associated with the Radon transform, see [20]; for $\alpha \geq 0$ we define the norm

$$\|g\|^2_{H^\alpha(Z)} = \int_0^\pi \|g(\cdot, \omega(\theta)\|^2_{H^\alpha(\mathbb{R})} d\theta$$

and therewith for each $\alpha \geq 0$ Sobolev spaces on cylinders can be defined by

$$H^\alpha(Z) = \{g \in L_2(Z) : \|g\|_{H^\alpha(Z)} < \infty\} \ .$$

Within this topology one has $R : H^\alpha(\mathbb{R}^2) \to H^{\alpha+\frac{1}{2}}(Z)$ (in two dimensions). Moreover, it holds $R^*R : H^\alpha(\mathbb{R}^2) \to H^{\alpha+1}(\mathbb{R}^2)$, i.e. R^*R is of order $t = -1/2$. Since for noisy data we have at most $g^\delta \in L_2(Z)$, the right hand side R^*g^δ belongs to $H^{1/2}(\mathbb{R}^2)$ leading in the notion $R^*R : H^{t+\sigma}(\mathbb{R}^2) \to H^{-t+\sigma}(\mathbb{R}^2)$ to $-t + \sigma = 1/2$ and consequently to $\sigma = 0$. This, however, does not assure optimal complexity (which is guaranteed at least for $\sigma > 1$). To circumvent this misfortune, it is reasonable to assume the exact data R^*g to be of higher Sobolev smoothness. To arrive although at the realistic measurement situation, we suggest to involve the so-called Sobolev embedding, as e.g. considered in [26, 27],

$$id_\tau R^*g + noise = R^*g^\delta \in H^{1/2}(\mathbb{R}^2) \ ,$$

where for arbitrary $s \in \mathbb{R}$

$$id_\tau : H^s(\mathbb{R}^2) \hookrightarrow H^{s-\tau}(\mathbb{R}^2) \ .$$

Therefore, instead of (51), we may consider the problem

$$(id_\tau \circ R^*R)f = R^*g^\delta \ . \tag{52}$$

Note that the incoporation of the Sobolev embedding operator may increase the overall ill-posedness of the reconstruction problem. But nevertheless, it provides by reversing the operator order a binding screw for theoretically fine tuning the procedure towards optimal performance.

In particular, equation (52) yields in the notion $(id_\tau \circ R^*R) : H^{t+\sigma}(\mathbb{R}^2) \to H^{-t+\sigma}(\mathbb{R}^2)$ a new operator order, namely

$$t = \frac{\tau - 1}{2} \ .$$

Due to the constraint $R^*g^\delta \in H^{1/2}(\mathbb{R}^2)$, it follows that $\sigma = 1/2 + t$. For the purpose of enabling $s^* > 0$ (in two dimensions), we have to restrict the choice of τ to

$$\sigma = \frac{1}{2} + \frac{\tau - 1}{2} = \frac{\tau}{2} > 1 \ .$$

This in turn yields a Sobolev source condition on the solution of the form $f \in H^{\tau-1/2}(\mathbb{R}^2)$. In order to finally ensure $s^* > 0$, one has to choose the regularity bounds $\gamma, \tilde{\gamma}$ and vanishing moments $d, \tilde{d}$ of the wavelet systems large enough (which can be done without further difficulties).

We wish to remark, that in the presented numerical experiment we have used for the reason of simplicity just Haar wavelet systems. Even in this situation, the numerical results of the proposed adaptive scheme are already considerably sparse compared to the solution of the full system. Moreover, the usage of Haar wavelets (or B-splines in general) allows an exact computation of stiffness matrix entries avoiding additional numerical errors as usually introduced by quadrature rules.

In particular, in our approach we have chosen a Haar wavelet frame $\Psi = \{\psi_\lambda\}_{\lambda\in\Lambda}$ with bounds A, B and have transformed the function space equation (52) into an infinite dimensional ℓ_2 system, see formula (12),

$$\boldsymbol{S}\boldsymbol{f} = \boldsymbol{g}^\delta, \tag{53}$$

where $\boldsymbol{S} = \boldsymbol{D}^{-t}F(id_\tau \circ R^*R)F^*\boldsymbol{D}^{-t}$ is the stiffness matrix of $(id_\tau \circ R^*R)$ preconditioned by a matrix $\boldsymbol{D}^{-t}$, $F^*\boldsymbol{D}^{-t}\boldsymbol{f} = f$, and $\boldsymbol{g}^\delta = \boldsymbol{D}^{-t}FR^*g^\delta$ is the preconditioned right hand side. Wavelet based preconditioning is a widely used tool, which allows to speed up significantly the computation of adaptive methods, cf. [9, 24]. In our case we use the wavelet-based preconditioner $\boldsymbol{D}^{-t}$ with $\boldsymbol{D}_{\lambda,\lambda'} = 2^{|\lambda|}\delta_{\lambda,\lambda'}$, where $2t = (1-d)/2$ is the order of the Radon operator R.

The resulting inexact (adaptive) iteration then reads as

$$\begin{aligned}
\tilde{\boldsymbol{f}}^\delta_{m+1} &= \boldsymbol{APPROX}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta] \\
&:= \boldsymbol{COARSE}_{r^\delta(m)}\big[\tilde{\boldsymbol{f}}^\delta_m - \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\big],
\end{aligned} \tag{54}$$

with accuracy/refinement map $r^\delta(m)$ and truncation index m_* chosen by the adaptive discrepancy principle (D). By $\tilde{\boldsymbol{f}}^\delta$ we denote the result of the truncated iteration (54) for given noise level δ. A corresponding approximation $\tilde{f}^\delta$ of the continuous solution f can be computed from the preconditioned iteration (54) by $\tilde{f}^\delta = F^*\boldsymbol{D}^{-t}\tilde{\boldsymbol{f}}^\delta$.

For the numerical experiments we have used the so-called *Shepp-Logan Phantom* f and the associated Radon transformed (noise free) data g, see Figure 1. As the main goal is to reduce the computational complexity, we illustrate the performance of the algorithm for a slightly modified truncation rule, namely $C_{r^\delta(m-1)+j}(\tilde{\boldsymbol{f}}^\delta_m) \sim \delta$. This rule yields also a convergent numerical scheme but delivers a sparser approximation of the solution of the inverse problem while having a remarkable better rate of the best N-term approximation as it is the case for the $C_{r^\delta(m-1)+j}(\tilde{\boldsymbol{f}}^\delta_m) \sim \delta^2$ truncation criterion (see below a more precise reasoning). In Figure 4, different approximations $\tilde{f}^\delta$ for different noisy right hand sides g^δ are illustrated. The final frame grids of the individual reconstructions are given in Figure 5. All the reconstructions were done for different relative noise levels $\delta_{rel} = \|g^\delta - g\|/\|g\|$.

In order to verify the effectiveness of the proposed scheme, we compare in Figure 2 the convergence rates as well as the computational complexity of the nonlinear approximation scheme (our proposed adaptive schemes (54)) with a linear approximation method. For our comparison,

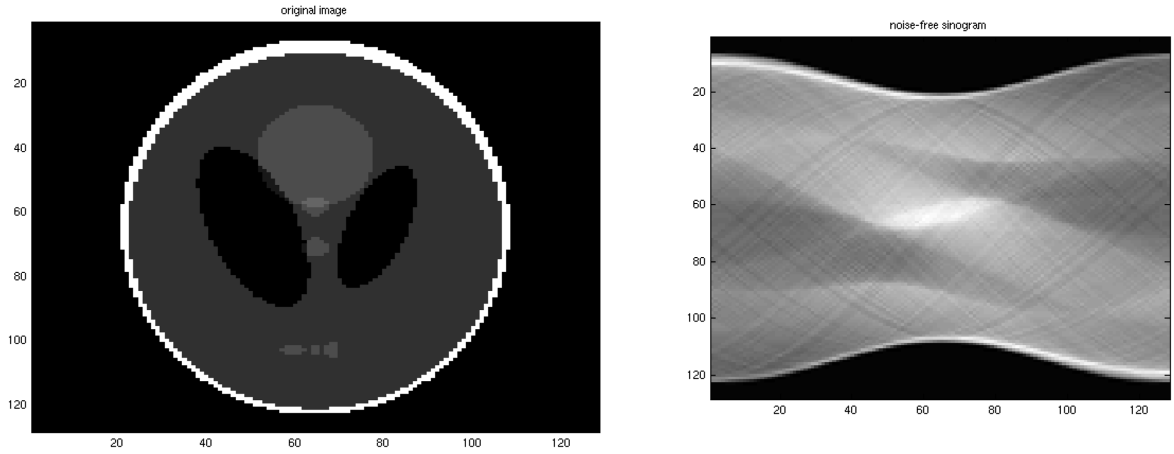

Figure 1: Left: Shepp-Logan Phantom image, right: associated noise free sinogram.

the linear approximation method was obtained by carrying out all iterations of the Landweber method with the full system matrix. As one can observe, for $\delta \to 0$, the convergence rates for $\tilde{f}^\delta \to f$ of the adaptive and non-adaptive approximations are nearly the same. The preconditioning step does not effect the convergence rates. However, considering the computational complexity (which is here defined as the total number of floating point operations), it can be observed that adaptivity significantly reduces the number of operations (compared to the linear approximation scheme) that are required to compute a good approximation to the solution of the inverse problem. The gain of nonlinear approximation increases as $\delta \to 0$, see Figure 2. This effect can be explained by the fact that for small δ the resolution of the solution becomes in principle finer and finer. Therefore, an adaptive choice of relevant coefficients has much more impact than representing the solution with respect to all possible coefficients (as usually done by linear approximation schemes). One the other hand, the sparsity of the representation of the reconstructed solution becomes for $\delta \to 0$ smaller, see Figure 3. But this is a natural consequence of the defined iteration scheme (54). In particular, this becomes clear due to the truncation condition (45) of the adaptive discrepancy principle (D), in which the accuracy term $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m)$ is compared with the squared noise level δ^2 (the term $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m)$ essentially involves of the accuracy tolerance $\varepsilon_{r^\delta(m)}$ of the routine $\boldsymbol{APPROX}$ implying that truncation accuracy also tends to zero as $\delta \to 0$). As an very interesting numerical observation, comparing the truncation rule $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m)$ with δ instead of δ^2, yields considerably sparser representations of the solution as well as an increased rate of the best N-term approximation. Unfortunately, the convergence order decreases from $\delta^{2/3}$ for $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m) \sim \delta^2$ to $\delta^{1/3}$ for $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m) \sim \delta$, see Figure 3.

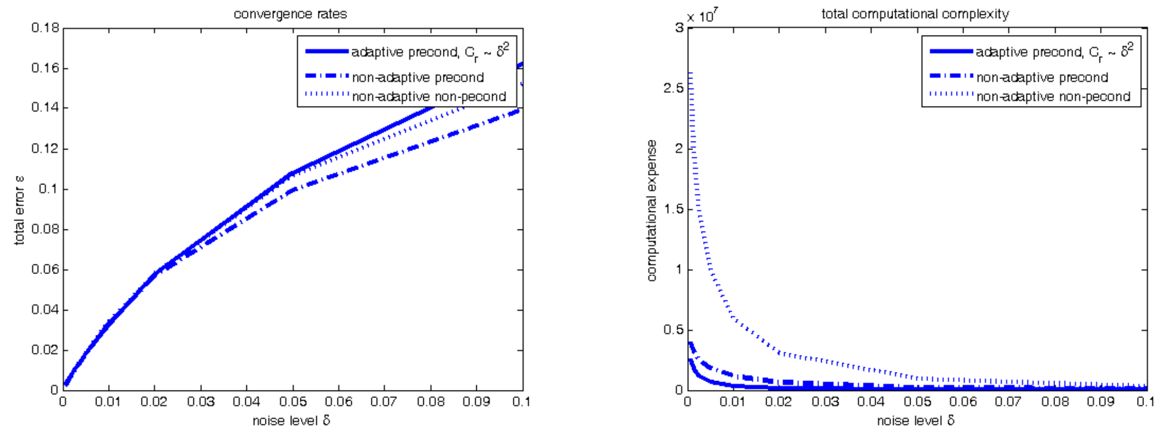

Figure 2:
Left: total approximation error vs. noise level, right: total computational complexity vs. noise level for 1. adaptive preconditioned iteration, 2. non-adaptive preconditioned iteration and 3. non-adaptive iteration without preconditioning.

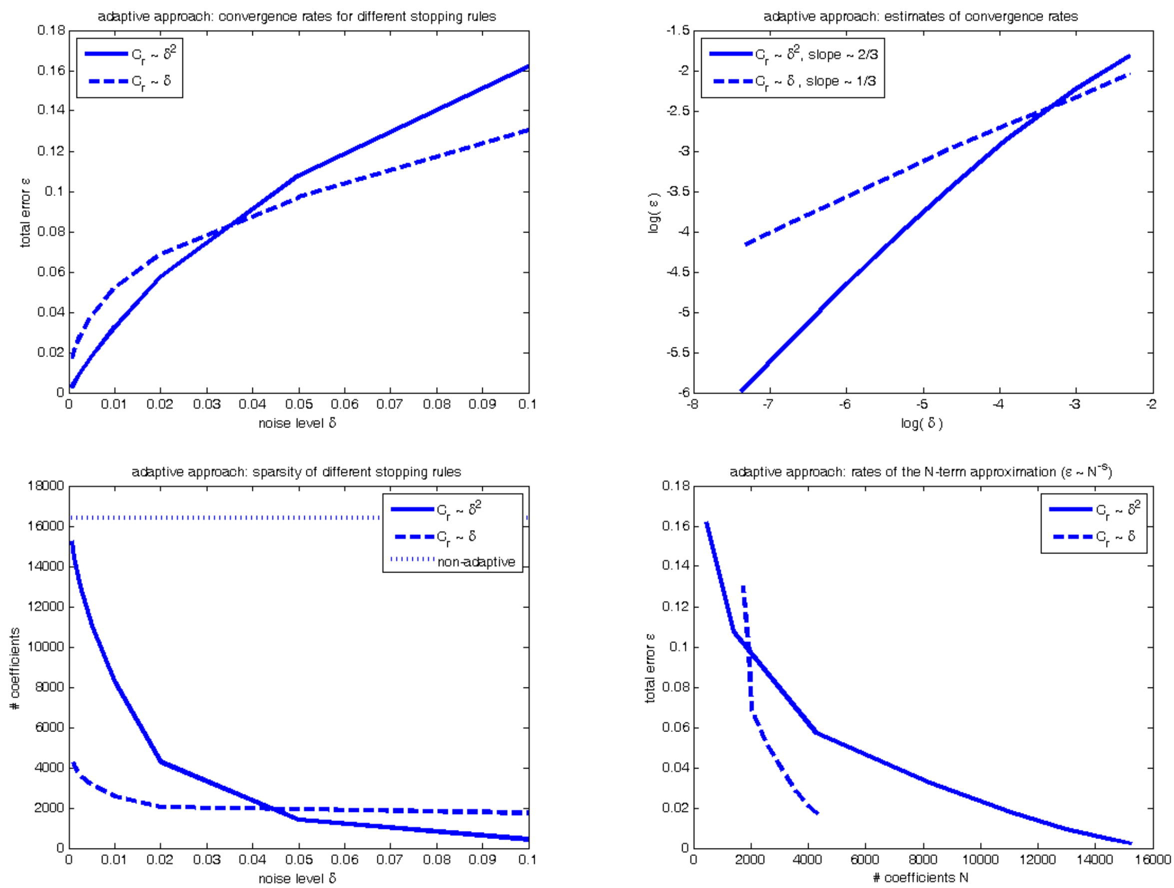

Figure 3:
Comparison of the truncation rules $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}_m^\delta) \sim \delta^2$ and $C_{r^\delta(m)}(\tilde{\boldsymbol{f}}_m^\delta) \sim \delta$. From top left to bottom right: total error ε vs. noise level δ, estimates of the convergence order, number of the non-zero coefficients at truncation step vs. noise level, total error vs. nonzero coefficients number.

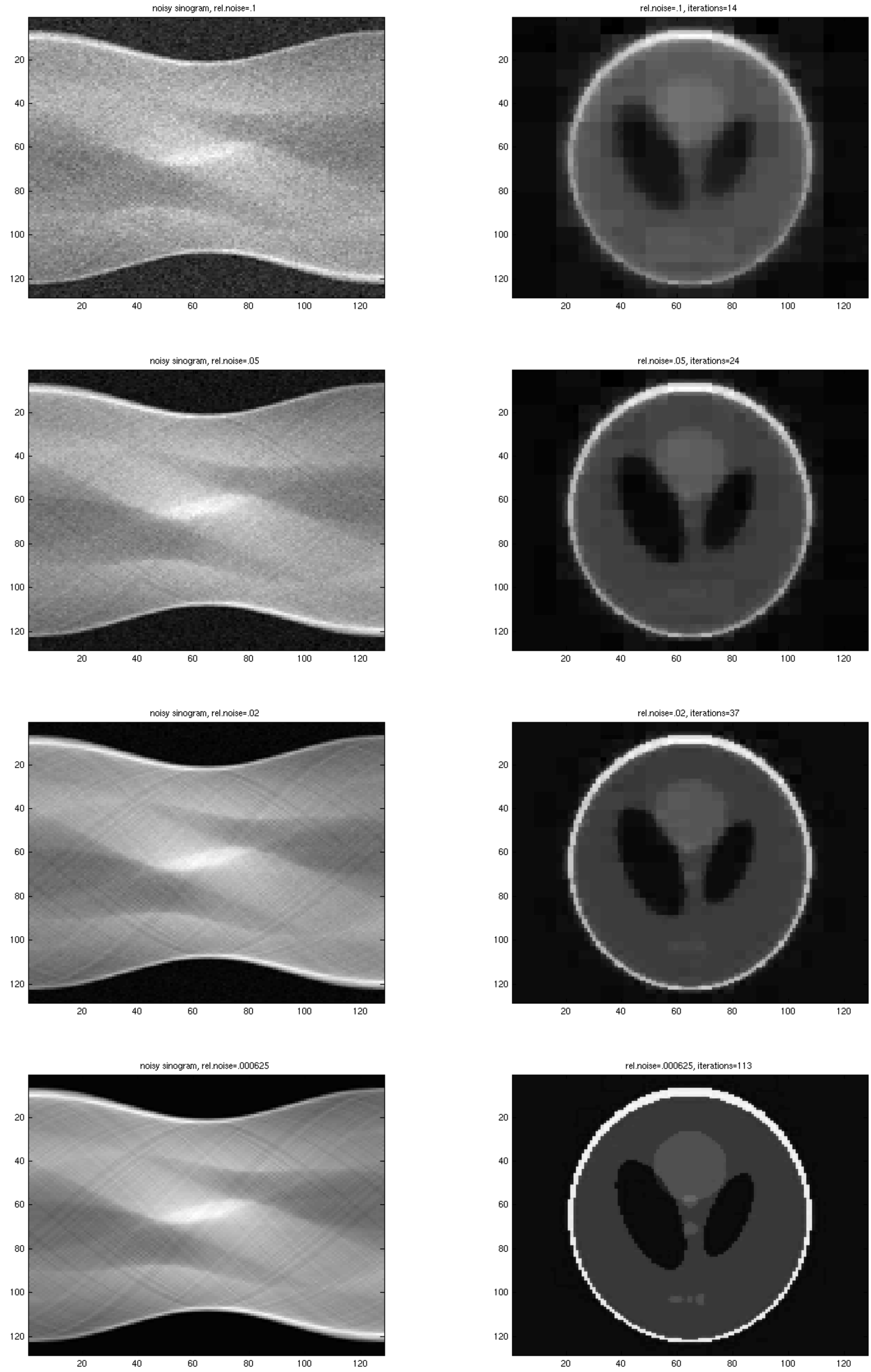

Figure 4: Adaptive Landweber approximations with modified truncation rule $C_{r^{\delta}(m)}(\tilde{\boldsymbol{f}}_m^{\delta}) \sim \delta$ for different noise levels $\delta_{rel} = 10\%, 5\%, 2\%, 0.00625\%$. The left column shows the noisy data g^{δ}; the right column shows the reconstructions.

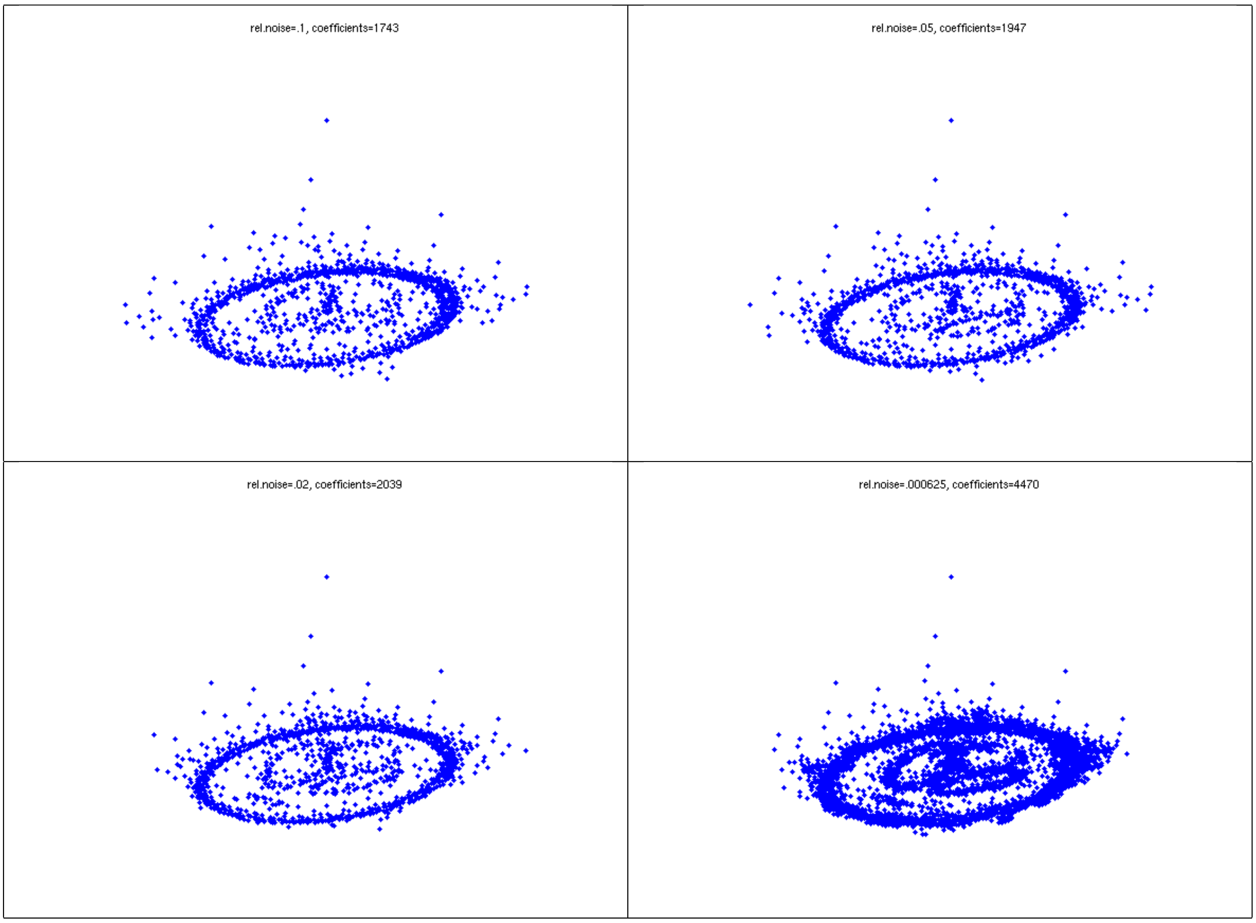

Figure 5: Different resulting wavelet frame grids associated to different adaptive reconstructions for noise levels $\delta_{rel} = 10\%, 5\%, 2\%, 0.00625\%$ shown in Figure 4. The individual sub-figures are three-dimensional: the layer ordering is from coarse to fine scale frame coefficients (from top to bottom). Within each layer we have visualized the label/location of the frame coefficients (not its magnitude).

Appendix

Proof of Lemma 4

At first, we observe that

$$\|\tilde{\boldsymbol{f}}^\delta_{m+1} - \boldsymbol{f}^\dagger\|^2 - \|\tilde{\boldsymbol{f}}^\delta_m - \boldsymbol{f}^\dagger\|^2 = \langle \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m, \tilde{\boldsymbol{f}}^\delta_{m+1} + \tilde{\boldsymbol{f}}^\delta_m - 2\boldsymbol{f}^\dagger\rangle$$

$$\begin{aligned}
&= \langle \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \tilde{\boldsymbol{f}}^\delta_{m+1} + \tilde{\boldsymbol{f}}^\delta_m - 2\boldsymbol{f}^\dagger\rangle \\
&\quad + \langle -\beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \tilde{\boldsymbol{f}}^\delta_{m+1} + \tilde{\boldsymbol{f}}^\delta_m - 2\boldsymbol{f}^\dagger\rangle \\
&= \langle \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \\
&\qquad \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\rangle \\
&\quad + \langle \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], 2\tilde{\boldsymbol{f}}^\delta_m - 2\boldsymbol{f}^\dagger\rangle \\
&\quad + 2\langle -\beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \\
&\qquad \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\rangle \\
&\quad + \langle -\beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \\
&\qquad -\beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta] + 2\tilde{\boldsymbol{f}}^\delta_m - 2\boldsymbol{f}^\dagger\rangle \\
&\le (\varepsilon^C_{r^\delta(m)})^2 + 2\varepsilon^C_{r^\delta(m)}(\|\tilde{\boldsymbol{f}}^\delta_m\| + \|\boldsymbol{f}^\dagger\|) \\
&\quad + 2\langle \beta F L^* L F^* \tilde{\boldsymbol{f}}^\delta_m - \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - \beta F L^* g^\delta, \\
&\qquad \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\rangle \\
&\quad + 2\langle -\beta F L^* L F^* \tilde{\boldsymbol{f}}^\delta_m + \beta F L^* g^\delta, \tilde{\boldsymbol{f}}^\delta_{m+1} - \tilde{\boldsymbol{f}}^\delta_m + \beta \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \beta \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\rangle \\
&\quad + \underbrace{\beta^2 \|\boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \boldsymbol{RHS}_{r^\delta(m)}[g^\delta]\|^2}_{=:T_1} \\
&\quad + \underbrace{2\beta\langle -\boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \tilde{\boldsymbol{f}}^\delta_m - \boldsymbol{f}^\dagger\rangle}_{=:T_2} \\
&\le (\varepsilon^C_{r^\delta(m)})^2 + 2\varepsilon^C_{r^\delta(m)}(\|\tilde{\boldsymbol{f}}^\delta_m\| + \|\boldsymbol{f}^\dagger\|) + \beta(\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)})\varepsilon^C_{r^\delta(m)} \\
&\quad + \beta\|F L^*\|\|L\tilde{f}^\delta_m - g^\delta\|\varepsilon^C_{r^\delta(m)} + T_1 + T_2 \\
&\le (\varepsilon^C_{r^\delta(m)})^2 + 2\varepsilon^C_{r^\delta(m)}(\|\tilde{\boldsymbol{f}}^\delta_m\| + \|\boldsymbol{f}^\dagger\|) + \beta(\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)})\varepsilon^C_{r^\delta(m)} \\
&\quad + \frac{1}{2}\big(\beta^2\|\boldsymbol{S}\|\|L\tilde{f}^\delta_m - g^\delta\|^2 + (\varepsilon^C_{r^\delta(m)})^2\big) + T_1 + T_2 \quad ,
\end{aligned} \tag{55}$$

The quantities T_1 and T_2 can be estimated as follows

$$\begin{aligned}
T_1 &= \beta^2\|\boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m]\| \\
&\le \beta^2\big(\|\boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - F L^* g^\delta - \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] + \boldsymbol{S}\tilde{\boldsymbol{f}}^\delta_m\| + \|F L^* g^\delta - \boldsymbol{S}\tilde{\boldsymbol{f}}^\delta_m\|\big)^2 \\
&\le 2\beta^2\big((\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})^2 + \|\boldsymbol{S}\|\|g^\delta - L\tilde{f}^\delta_m\|^2\big)
\end{aligned} \tag{56}$$

and

$$
\begin{aligned}
T_2 &= 2\beta\big[\langle \tilde{\boldsymbol{f}}^\delta_m, \boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m]\rangle \\
&\qquad - \langle \boldsymbol{f}^\dagger, \boldsymbol{RHS}_{r^\delta(m)}[g^\delta] - \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m]\rangle\big] \\
&= 2\beta\big[-\|g^\delta\|^2 + 2\langle \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \tilde{\boldsymbol{f}}^\delta_m\rangle - \langle \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m], \tilde{\boldsymbol{f}}^\delta_m\rangle \\
&\qquad + \|g^\delta\|^2 + \langle FL^*g^\delta - \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \tilde{\boldsymbol{f}}^\delta_m\rangle - \langle FL^*g^\delta, \tilde{\boldsymbol{f}}^\delta_m\rangle \\
&\qquad + \langle FL^*g^\delta - \boldsymbol{RHS}_{r^\delta(m)}[g^\delta], \boldsymbol{f}^\dagger\rangle - \langle FL^*g^\delta, \boldsymbol{f}^\dagger\rangle \\
&\qquad + \langle \boldsymbol{APPLY}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m] - \boldsymbol{S}\tilde{\boldsymbol{f}}^\delta_m, \boldsymbol{f}^\dagger\rangle + \langle \boldsymbol{S}\tilde{\boldsymbol{f}}^\delta_m, \boldsymbol{f}^\dagger\rangle\big] \\
&\le 2\beta\big[-(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + \|g^\delta\|^2 + \varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| - \langle g^\delta, LF^*\tilde{\boldsymbol{f}}^\delta_m\rangle \\
&\qquad + \varepsilon^R_{r^\delta(m)}\|\boldsymbol{f}^\dagger\| - \langle g^\delta, LF^*\boldsymbol{f}^\dagger\rangle + \varepsilon^A_{r^\delta(m)}\|\boldsymbol{f}^\dagger\| + \langle LF^*\tilde{\boldsymbol{f}}^\delta_m, LF^*\boldsymbol{f}^\dagger\rangle\big] \\
&= 2\beta\big[-(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + \varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\| \\
&\qquad + \|g^\delta\|^2 - \langle g^\delta, LF^*\tilde{\boldsymbol{f}}^\delta_m\rangle - \langle g^\delta, g\rangle + \langle LF^*\tilde{\boldsymbol{f}}^\delta_m, g\rangle\big] \\
&= 2\beta\big[-(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + \varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\| \\
&\qquad + \langle g^\delta - g, g^\delta - LF^*\tilde{\boldsymbol{f}}^\delta_m\rangle\big] \\
&\le 2\beta\big[-(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + \varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\| \\
&\qquad + \|g^\delta - g\|\|g^\delta - LF^*\tilde{\boldsymbol{f}}^\delta_m\|\big] \\
&\le 2\beta\big[-(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + \varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\|\big] \\
&\qquad + \beta(\delta^2 + \|g^\delta - L\tilde{f}^\delta_m\|^2). \qquad (57)
\end{aligned}
$$

Inserting the estimates (56) for T_1 and (57) for T_2 into (55) finally yields

$$
\begin{aligned}
&\|\tilde{\boldsymbol{f}}^\delta_{m+1} - \boldsymbol{f}^\dagger\|^2 - \|\tilde{\boldsymbol{f}}^\delta_m - \boldsymbol{f}^\dagger\|^2 \\
&\le \frac{3}{2}(\varepsilon^C_{r^\delta(m)})^2 + 2\varepsilon^C_{r^\delta(m)}(\|\tilde{\boldsymbol{f}}^\delta_m\| + \|\boldsymbol{f}^\dagger\|) + \beta(\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)})\varepsilon^C_{r^\delta(m)} \\
&\qquad + \frac{1}{2}\beta^2\|\boldsymbol{S}\|\|L\tilde{f}^\delta_m - g^\delta\|^2 \\
&\qquad + 2\beta^2(\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})^2 + 2\beta^2\|\boldsymbol{S}\|\|g^\delta - L\tilde{f}^\delta_m\|^2 \\
&\qquad - 2\beta(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + 2\beta\big(\varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\|\big) \\
&\qquad + \beta\delta^2 + \beta\|g^\delta - L\tilde{f}^\delta_m\|^2 \\
&= \frac{3}{2}(\varepsilon^C_{r^\delta(m)})^2 + 2\varepsilon^C_{r^\delta(m)}(\|\tilde{\boldsymbol{f}}^\delta_m\| + \|\boldsymbol{f}^\dagger\|) + \beta(\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)})\varepsilon^C_{r^\delta(m)} \\
&\qquad + 2\beta^2(\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})^2 + 2\beta\big(\varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\|\big) \\
&\qquad + \beta\delta^2 - 2\beta(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 + \beta\big(1 + \frac{3}{2}\beta\|\boldsymbol{S}\|\big)\|g^\delta - L\tilde{f}^\delta_m\|^2 \\
&\le \beta C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m) + \beta\delta^2 + \beta\big(\frac{3}{2}\beta\|\boldsymbol{S}\| - 1\big)(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2, \qquad (58)
\end{aligned}
$$

where we have introduced for ease of notation

$$
\begin{aligned}
C_{r^\delta(m)}(\tilde{\boldsymbol{f}}^\delta_m) \;:=\; & \frac{3}{2\beta}(\varepsilon^C_{r^\delta(m)})^2 + \frac{2}{\beta}\varepsilon^C_{r^\delta(m)}(\|\tilde{\boldsymbol{f}}^\delta_m\| + \|\boldsymbol{f}^\dagger\|) + (\varepsilon^A_{r^\delta(m)} + \varepsilon^R_{r^\delta(m)})\varepsilon^C_{r^\delta(m)} \\
& + 2\beta(\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})^2 + 2\big(\varepsilon^R_{r^\delta(m)}\|\tilde{\boldsymbol{f}}^\delta_m\| + (\varepsilon^R_{r^\delta(m)} + \varepsilon^A_{r^\delta(m)})\|\boldsymbol{f}^\dagger\|\big) \\
& + \big(1 + \frac{3}{2}\beta\|\boldsymbol{S}\|\big)(\varepsilon^A_{r^\delta(m)} + 2\varepsilon^R_{r^\delta(m)})\|\tilde{\boldsymbol{f}}^\delta_m\| \; . \qquad (59)
\end{aligned}
$$

Since (19) was assumed, and thanks to (58), monotony can be deduced

$$
\|\tilde{\boldsymbol{f}}^\delta_{m+1} - \boldsymbol{f}^\dagger\|^2 - \|\tilde{\boldsymbol{f}}^\delta_m - \boldsymbol{f}^\dagger\|^2 < \beta\big(\frac{3}{2}\beta\|\boldsymbol{S}\| - 1\big)(1-c)(\boldsymbol{RES}_{r^\delta(m)}[\tilde{\boldsymbol{f}}^\delta_m, g^\delta])^2 < 0 \; . \qquad (60)
$$

Acknowledgement

The authors thank S. Dahlke and T. Raasch for encouraging discussions on nonlinear and adaptive approximation and for providing an implementation guideline. The work of R. Ramlau was supported by the Austrian Science Fund (FWF), Grants FWF-P19029-N18, FWF-P19496-N18 and FWF-P20237-N14. G. Teschke gratefully acknowledges the partial support by Deutsche Forschungsgemeinschaft Grants TE 354/1-2, TE 354/3-1, TE 354/4-1, TE 354/5-1. M. Zhariy gratefully acknowledges the partial support by Deutsche Forschungsgemeinschaft Grant TE 354/3-1.

References

[1] S. Anthoine. Different wavelet-based approaches for the separation of noisy and blurred mixtures of components. application to astrophysical data. *PhD Thesis*, 2005.

[2] F. Bauer and T. Hohage. A lepskij-type stopping rule for regularized newton methods. *Inverse Problems*, 21:1975–1991, 2005.

[3] T. Bonesky and P. Maass. Iterated soft shrinkage with adaptive operator evaluations. *Journal of Inverse and Ill-Posed Problems*, to appear, 2008.

[4] K. Bredies, D.A. Lorenz, and P. Maass. A generalized conditional gradient method and its connection to an iterative shrinkage method. *to appear in: Computational Optimization and Application*, 2005.

[5] O. Christensen. *An Introduction to Frames and Riesz Bases.* Birkhäuser, Boston, 2003.

[6] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods ii - beyond the elliptic case. *Found. Comput. Math.*, 2:203–245, 2002.

[7] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Adaptive wavelet methods for elliptic operator equations — Convergence rates. *Mathematics of Computation*, 70(233):27–75, 2001.

[8] S. Dahlke and P. Maass. An outline of adaptive wavelet galerkin methods for tikhonov regularization of inverse parabolic problems. *In: Recent Development in Theory & Numerics, Proceedings of the International Conference on Inverse Problems, World Scientific*, pages 56–66, 2002.

[9] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numerica, Cambridge University Press*, 6:55–228, 1997.

[10] I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57:1413–1541, 2004.

[11] I. Daubechies, M. Fornasier, and I. Loris. Accelerated projected gradient methods for linear inverse problems with sparsity constraints. *J. Fourier Anal. Appl.*, to appear, 2008.

[12] I. Daubechies and G. Teschke. Wavelet–based image decomposition by variational functionals. *Proc. SPIE Vol. 5266, p. 94-105, Wavelet Applications in Industrial Processing; Frederic Truchetet; Ed.*, Feb. 2004.

[13] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Applied and Computational Harmonic Analysis*, 19(1):1–16, 2005.

[14] I. Daubechies, G. Teschke, and L. Vese. Iteratively solving linear inverse problems with general convex constraints. *Inverse Problems and Imaging*, 1(1):29–46, 2007.

[15] I. Daubechies, G. Teschke, and L. Vese. On some iterative concepts for image restoration. *Advances in Imaging and Electron Physics*, 150, 2008.

[16] M. Defrise and C. De Mol. A note on stopping rules for iterative methods and filtered svd. In P. C. Sabatier, editor, *Inverse Problems: An Interdisciplinary Study*, pages 261–268. Academic Press, 1987.

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

[18] M. Fornasier. Domain decomposition methods for linear inverse problems with sparsity constraints. *Inverse Problems*, 23:2505, 2007.

[19] H. Gfrerer. An a posteriori parameter choice for ordinary and iterated tikhonov regularization of ill posed problems leading to optimal convergence rates. *Math. Comp.*, 49:507–522, 1987.

[20] A. K. Louis. *Inverse und schlecht gestellte Probleme.* Teubner, Stuttgart, 1989.

[21] P. Mathe. The lepskii principle revisited. *Inverse Problems*, 21:L11–L15, 2005.

[22] V.A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[23] F. Natterer. *The Mathematics of Computerized Tomography.* Teubner, Stuttgart, 1986.

[24] T. Raasch. Adaptive wavelet and frame schemes for elliptic and parabolic equations. *PhD Thesis*, 2007.

[25] R. Ramlau. A modified Landweber–method for inverse problems. *Journal for Numerical Functional Analysis and Optimization*, 20(1/2):79–98, 1999.

[26] R. Ramlau and G. Teschke. Regularization of Sobolev Embedding Operators and Applications Part I: Fourier and Wavelet based Methods. *Sampling Theory in Signal and Image Processing*, 3(2):175–196, 2004.

[27] R. Ramlau and G. Teschke. Regularization of Sobolev Embedding Operators and Applications Part II: Data Driven Regularization and Applications. *Sampling Theory in Signal and Image Processing*, 3(3):225–246, 2004.

[28] R. Ramlau and G. Teschke. A Projection Iteration for Nonlinear Operator Equations with Sparsity Constraints. *Numerische Mathematik*, 104:177–203, 2006.

[29] R. Stevenson. Adaptive solution of operator equations using wavelet frames. *SIAM J. Numer. Anal.*, (41):1074–1100, 2003.

[30] R. Stevenson. On the compressibility of operators in wavelet coordinates. *SIAM Journal on Mathematical Analysis*, 35(5):1110–1132, 2004.

[31] U. Tautenhahn and Hämarik. The use of monotonicity for choosing the regularization parameter in ill -posed problems. *Inverse Problems*, 15:1487–1505, 1999.

[32] G. Teschke. Multi-Frame Representations in Linear Inverse Problems with Mixed Multi-Constraints. *Applied and Computational Harmonic Analysis*, 22:43 – 60, 2007.

[33] G.M. Vainikko. The principle of the residual for a class of regularization methods. *USSR Comput. Math. Math. Phys.*, 22:1–19, 1982.

[34] M. Zhariy. Adaptive frame based regularization methods for solving linear ill-posed inverse problems. *PhD Thesis*, 2008.