Sparse Recovery in Inverse Problems

Ronny Ramlau and Gerd Teschke

Abstract. Within this chapter we present recent results on sparse recovery algorithms for inverse and ill-posed problems, i.e. we focus on those inverse problems in which we can assume that the solution has a sparse series expansion with respect to a preassigned basis or frame. The presented approaches to approximate solutions of inverse problems are limited to iterative strategies that essentially rely on the minimization of Tikhonov-like variational problems, where the sparsity constraint is integrated through ℓ_p norms. In addition to algorithmic and computational aspects, we also discuss in greater detail regularization properties that are required for cases in which the operator is ill-posed and no exact data are given. Such scenarios reflect realistic situations and manifest therefore its great suitability for "real-life" applications.

Keywords. inverse and ill-posed problems, regularization theory, convergence rates, sparse recovery, iterated soft-shrinkage, accelerated steepest descent, nonlinear approximation.

AMS classification. 41A46, 45Q05, 47A52, 49M99, 52A99, 65B99, 65K10.

1 Introduction

The aim of this chapter is to present technologies for the recovery of sparse signals in situations in which the given data are linked to the signal to be reconstructed through an ill-posed measurement model. In such scenarios one is typically faced with regularization issues and the construction of suitable methods that allow a stable reconstruction of sparse signals.

1.1 Road map of the chapter

Nowadays there exist a great variety of schemes realizing sparse reconstructions. Most of them are well-suited for finite or infinite dimensional problems but where the underlying physical model is well-posed. More delicate are those cases in which ill-posed operators are involved. So far, for linear ill-posed problems, there are numerous schemes available that perform quite well sparse reconstructions, e.g. [2, 14, 18, 19, 24, 25, 52]. The majority of these approaches rely on iterative concepts in which adequate sparsity constraints are involved. Within this chapter we do not discuss the pros and cons of all these methods. In the context of linear problems we just concentrate on one new approach that involves a complete different but very powerful technology - that is adaptive approximation. The second focus of this chapter is on the generalization of conventional iterative strategies to nonlinear ill-posed problems.

Therefore the road map for this chapter is as follows: In Section 2 we collect the basics on inverse problems. To elaborate the differences between well- and ill-posedness and the concepts of regularization theory as simple as possible we limit ourselves in this introductory section to linear problems. After this preliminary part we continue in Section 3 with linear problems and present a sparse recovery principle that essentially relies on the theory of adaptive approximation. The main ingredient that ensures stable recovery are sophisticated refinement strategies. In Section 4 we turn then to nonlinear ill-posed problems and discuss in greater detail Tikhonov regularization with sparsity constraints. The established regularization properties include convergence results and convergence rates for a-priori as well as for a-posteriori parameter rules. After the general discussion on Tikhonov regularization we focus within the following Sections 5 and 6 on the development of implementable algorithms to numerically realize sparse recovery. The first method presented in Section 5 relies on the surrogate functional technology. This approach results in a Landweber-type iteration where a shrinkage operation is applied in each iteration step. This method can be generalized to general sparsity constraints, but fails to be numerically efficient. To overcome this deficiency, we introduce in Section 6 a slightly modified concept leading to a very similar iteration, but where in each iteration a projection on a preassigned ℓ_1 ball is applied. Moreover, this new iteration is designed with an adaptive step length control resulting in a numerically very efficient method.

1.2 Remarks on sparse recovery algorithms

As mentioned above, we discuss in this chapter two different species of sparse recovery algorithms. The first species developed for linear inverse problems relies on nonlinear approximation, the second species designed for nonlinear inverse problems relies on linear approximation.

In principle, when it comes to numerical realizations, we are faced with the problem that we can only treat finite index sets. Therefore one has to answer the question which coefficients should be involved in the reconstruction process and which can be neglected. Linear approximation simply suggests a truncation of the infinite index set. In a wavelet framework this would mean to limit the number resolution scales. For many problems in which the solution is supposed to have a certain Sobolev smoothness, this proceeding might yield reasonable results. Nevertheless, there are still cases in which linear approximation fails to yield optimal results. Then often nonlinear approximation concepts are much better suited. The reason why nonlinear strategies perform better than standard linear methods is due to the properties of the solution and the operator. To clarify this statement, we introduce by x_N the best N-term approximation of the solution x. Considering bases or frames of sufficiently smooth wavelet type (e.g. wavelets of order d), it is known 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, and x is in the Besov space $B_\tau^{sn+t}(L_\tau(\Omega))$ with $\tau = (1/2+s)^{-1}$, then

$$\sup_{N\in\mathbb{N}} N^s \|x - x_N\| < \infty \,.$$

The condition $x \in B_\tau^{sn+t}(L_\tau(\Omega))$ is much milder than requiring $x \in H^{sn+t}(\Omega)$ that would be needed to guarantee the same rate of convergence with linear approximation. 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. Therefore, special a-priori information about x and/or the operator is required. In certain cases, e.g. the tomographic reconstruction problem analyzed in [35], this information can be derived. A suitable model class for the tomographic reconstruction problem are piecewise constant functions with jumps along smooth manifolds. It is shown that such functions belong to the Sobolev space $H^{sd}(\Omega)$ with $sd < 1/2$. An adaptive approximation of such functions (when carried out in $L_2(\Omega)$) 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 [50, Rem. 4.3] 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 of this application, we therefore have that the solution x 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 and nonlinear approximation pays off. How nonlinear approximation strategies can be realized for linear inverse and ill-posed problems shall be discussed in great detail in Section 3. As Besov regularity directly translates into sparsity, the elaborated adaptive Landweber-type scheme performs a sparse recovery for x.

Sparse recovery algorithms for nonlinear inverse problems rely so far on linear approximation concepts that originate from the minimization of Tikhonov-like functionals. These concepts were originally developed for linear inverse problems within the last decade, see e.g. [2, 14, 18, 19, 24, 25, 52], and have led to many breakthroughs in a broad field of applications. They are due to its simple nature very easy to use and can be applied in various reformulations. The generalization of these methods to nonlinear problems has permitted an algorithmic realization of sparse recovery for problems that were by then not feasible. The main ingredients are a proper variational formulation of the data misfit term and an adequate involvement of the sparsity constraint either through an extra penalty term or an restriction of the possible solution set. These two concepts shall be elaborated in Sections 5 and 6 which are furnished with associated numerical experiments.

2 Classical Inverse Problems

In many applications in the natural sciences, medicine or imaging one has to determine the cause x of a measured effect y. A classical example is Computerized Tomography (CT), a medical application, where a patient is screened using x - rays. The observed damping of the rays is then used to reconstruct the density distribution of the body. In order to achieve such a reconstruction, the measured data and the searched for quantity have to be linked by a mathematical model, which we will denote by F (or A, if the model is linear). In an abstract setting, the determination of of the cause x can be stated as follows: Solve an operator equation

$$F(x) = y \ , \tag{2.1}$$

$F : X \to Y$, where X, Y are Banach (Hilbert) spaces. For the CT problem, the operator describing the connection between the measurements and the density distribution (in 2 dimensions) is given by the Radon transform,

$$y(s,\omega) = (Ax)(s,\omega) = \int\limits_{\mathbb{R}} x(s\omega + t\omega^{\perp})\, dt \ , \ s \in \mathbb{R}, \ \omega \in \mathcal{S}^1.$$

As in practice the observed data stems from measurements, one never has the exact data y available, but rather a noisy variant y^δ. In the following we might assume that at least a bound δ for the noise is available (e.g. if the accuracy of the measurement device is known):

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

In connection with Inverse Problems, the following questions arise:

(i) Does there exist a solution of equation (2.1) for given exact y?

(ii) Is the solution unique?

(iii) If the solution is determined from noisy data, how accurate is it?

(iv) How to solve (2.1)?

2.1 Preliminaries

In order to give a first idea on the problems that may be encountered for ill-posed problems, we will now consider a linear operator equation in finite dimensions. Assume $A \in \mathbb{R}^{n\times n}$, and we want to solve the linear system $Ax = y$ from noisy data y^δ. If we assume that A is invertible on $range(A)$ and also $y^\delta \in range(A)$ (which is already a severe restriction), then we can define

$$\begin{aligned} x^\dagger &:= A^{-1}y \\ x^\delta &:= A^{-1}y^\delta \ . \end{aligned}$$

With $x^\dagger - x^\delta = A^{-1}(y - y^\delta)$ the distance between x^δ and $x^\dagger$ can be estimated as follows,

$$\begin{aligned} \|x^\dagger - x^\delta\| &\leq \|A^{-1}\| \, \|y - y^\delta\| \\ &\leq \|A^{-1}\| \delta \ . \end{aligned} \tag{2.2}$$

If we additionally assume that A is symmetric and positive definite with $\|A\| \leq 1$, then A has an eigensystem (λ_i, x_i) with eigenvalues $0 < \lambda_i \leq 1$ and associated eigenvectors x_i. Moreover we have

$$\|A^{-1}\| = \frac{1}{\lambda_{min}} \Rightarrow \|x^\dagger - x^\delta\| \leq \frac{\delta}{\lambda_{min}}$$

Therefore, the reconstruction quality is of the same order $\mathcal{O}(\delta)$ as the data error, magnified only by the norm of the inverse operator. However, it turns out that $\mathcal{O}(\delta)$ estimates are only possible in a finite dimensional setting: Indeed, if we define the operator

$$Ax = \sum_{i=1}^{\infty} \lambda_i \langle x, x_i \rangle x_i$$

with orthonormal basis x_i and $\lambda_i \to 0$, then it is easily to see that the right hand side of estimate (2.2) explodes. In fact, for inverse problems with $dim\ range(A) = \infty$ and, e.g., compact operator, it is in general impossible to obtain convergence rates. Under additional assumptions on the solution of the equation $Ax = y$, the best possible convergence rate is given by

$$\|x - x^\delta\| = \mathcal{O}(\delta^s) \ , \quad s < 1. \tag{2.3}$$

The above considerations were based on the assumption $y^\delta \in range(A)$. As we will see in the following example, this is a severe restriction that will not hold in practice: Let us consider the integral equation

$$y(s) = \int_0^s x(t)dt \qquad 0 \le s \le 1.$$

If $x \in C^0[0,1]$, then it immediately follows that $y \in C^1[0,1]$ and

$$x(s) = y'(s), \qquad y(0) = 0 \ .$$

For noisy measurements this condition will not hold, as the noise will not only alter the initial value but also the smoothness of y^δ, as the data noise is usually not differentiable. The same also holds for Computerized Tomography: It can be shown [35] that the exact CT data belongs to the Sobolev space $H^{1/2}(\mathbb{R} \times \mathcal{S}^1)$, but for the noisy data we only have $y^\delta \in L_2$.

Now let us define *well-posed* and *ill-posed problems*.

Definition 2.1. Let $A : X \to Y$ linear operator and X, Y be topological spaces. Then the problem (A, X, Y) is well-posed if condition (i)-(iii) are fulfilled at the same time,

(i) $Ax = y$ has a solution for each $y \in Y$

(ii) the solution is unique

(iii) the solution depends continuously on the data, i.e.

$$y_n \to y, \ y_n = Ax_n, \Longrightarrow x_n \to x \text{ and } Ax = y \ .$$

If one of the conditions is violated, then the problem is ill posed.

Roughly speaking, well-posed problems allow for an error estimate as in (2.2), whereas the best possible rate for ill posed problems is as in (2.3).

Let us denote by $L(X, Y)$ the set of all linear and continuous operators $A : X \to Y$. An important class of operators that lead to ill-posed problems are *compact* operators.

Definition 2.2. An operator $A \in L(X, Y)$ is compact, if it maps bounded sets to relative compact sets. Or equivalently, for any bounded sequence $(x_n)_{n\in\mathbb{N}}$, the sequence $y_n = Ax_n$ has a convergent subsequence.

Integral operators are an important class of examples for compact operators.

Definition 2.3. Let $G \in \mathbb{R}^n$ be a bounded set and $k : G \times G \to G$. We define the integral operator K by

$$(Kx)(s) = \int_G k(s,t)x(t)dt \ .$$

Proposition 2.4. *Let $k \in C(G, G)$ and K be an integral operator considered between any of the spaces $L_2(G)$ and $C(G)$. Then K is compact. If $k \in L_2(G, G)$, then the integral operator $K : L_2(G) \to L_2(G)$ is compact.*

Another example for compact operators are Sobolev embedding operators. For bounded G and a real number $s > 0$, let us consider the map

$$i_s : H^s(G) \to L_2(G)\ , \quad \text{which is defined by } \ i_s x = x\ .$$

Here H^s denotes the standard Sobolev space. Then we have

Proposition 2.5. *The Sobolev embedding operator i_s is compact.*

Proposition 2.6. *Compact operators with $dim\ range(K) = \infty$ are not continuously invertible, i.e. they are ill-posed.*

Now let us assume that a given operator $A : H^s \to H^{s+t}$, $s \geq 0, t > 0$, is continuously invertible. As pointed out above, the measured data will not belong to H^{s+t} but rather to L_2. Therefore, we have to consider the operator equation between H^s and L_2, i.e. the equation $y = i_{s+t}(Ax)$. As a combination of a continuous and a compact operator, $i_s \circ A$ is also compact and therefore not continuously invertible - regardless of the invertibility of A.

A key ingredient for the stable inversion of compact operators is the spectral decomposition:

Proposition 2.7. *Let $K : X \to X$, X be a Hilbert space and assume that K is compact and self-adjoint (i.e, $\langle Kx, y \rangle = \langle x, Ky \rangle \ \forall x, y \in X$). By (λ_j, u_j) denote the set of eigenvalues λ_j and associated eigenvectors u_j with $Ku_j = \lambda_j u_j$. Then $\lambda_j \to 0$ $(if\ dim\ range(K) = \infty)$ and the functions u_j form an orthonormal basis of $\overline{range(K)}$ with*

$$Kx = \sum_{i=1}^{\infty} \lambda_i \langle x, u_i \rangle u_i\ .$$

The eigenvalue decomposition can be generalized to compact operators that are not self-adjoint. Let $K : X \to Y$ be given. The adjoint operator $K^* : Y \to X$ is formally defined by the equation

$$\langle Kx, y \rangle = \langle x, K^* y \rangle \quad \forall x, y\ .$$

We can then define the operator $K^*K : X \to X$ and find

$$\begin{aligned} \langle K^*Kx, y \rangle &= \langle Kx, Ky \rangle = \langle x, K^*Ky \rangle\ , \\ \langle K^*Kx, x \rangle &= \langle Kx, Kx \rangle = \|Kx\|^2\ , \end{aligned}$$

i.e., K^*K is selfadjoint and positive semi-definite, which also guarantees that all eigenvalues λ_i of K^*K are nonnegative. Therefore we have

$$K^*Kx = \sum_i \lambda_i \langle\, x, u_i \,\rangle u_i \ .$$

Defining

$$\begin{aligned} \sigma_i &= +\sqrt{\lambda_i} \\ Ku_i &= \sigma_i v_i \ , \end{aligned}$$

we find that the functions v_i also form an orthonormal system for X:

$$\begin{aligned} \langle\, v_i, v_j \,\rangle &= \frac{1}{\sigma_i \sigma_j} \langle\, Ku_i, Ku_j \,\rangle \\ &= \frac{1}{\sigma_i \sigma_j} \langle\, K^*Ku_i, u_j \,\rangle \\ &= \frac{\sigma_i}{\sigma_j} \langle\, u_i, u_j \,\rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \ , \end{aligned}$$

and get

$$\begin{aligned} Kx &= K(\sum_i \langle\, x, u_i \,\rangle u_i) = \sum_i \langle\, x, u_i \,\rangle K u_i = \sum_i \sigma_i \langle\, x, u_i \,\rangle v_i \ , \\ K^*y &= \sum_i \sigma_i \langle\, y, v_i \,\rangle u_i \ . \end{aligned}$$

The above decomposition of K is called the *singular value decomposition* and $\{\sigma_i, x_i, y_i\}$ is the singular system of K. The *generalized inverse* of K is defined as follows:

Definition 2.8. The generalized inverse $K^\dagger$ of K is defined as

$$\begin{aligned} dom(K^\dagger) &= range(K) \oplus range(K)^\perp \\ K^\dagger y &:= x^\dagger \\ x^\dagger &= \arg\min_x \|y - Kx\| \ . \end{aligned}$$

If the minimizer $x^\dagger$ of the functional $\|y - Kx\|^2$ is not unique then the one with minimal norm is taken.

Proposition 2.9. *The generalized solution $x^\dagger$ has the following properties*

(i) *$x^\dagger$ is the unique solution of $K^*Kx = K^*y$,*

(ii) *$Kx^\dagger = P_{R(K)}y$, where $P_{R(K)}$ denotes the orthogonal projection on the range of K,*

(iii) $x^\dagger$ *can be decomposed w.r.t. the singular system as*

$$x^\dagger = \sum_i \frac{1}{\sigma_i} \langle y, v_i \rangle u_i \; , \tag{2.4}$$

(iv) *the generalized inverse is continuous if and only if* $range(K)$ *is closed.*

A direct consequence of the above given representation of $x^\dagger$ is the so-called Picard condition:

$$y \in range(K) \Leftrightarrow \sum_i \frac{|\langle\, y, v_i \,\rangle|^2}{\sigma_i^2} < \infty \; .$$

The condition states that the moments of the right hand side y (w.r.t. to the system $\{v_i\}$) have to tend to zero fast enough in order to compensate the growth of $1/\sigma_i$.

What happens if we apply noisy data to formula (2.4)? Assume $y \in range(K)$, $y = Kx^\dagger$, and $y_l^\delta = y + \delta v_l$. Then for all l

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

but with

$$x^\delta = \sum_i \frac{1}{\sigma_i} \langle y_l^\delta, v_i \rangle u_i$$

we obtain

$$\|x - x^\delta\|^2 = \sum_i \frac{\delta^2}{\sigma_i^2} |\langle\, v_l, v_i \,\rangle|^2 = \frac{\delta^2}{\sigma_l^2} \to \infty \text{ as } l \to \infty \; ,$$

which shows that the reconstruction error can be arbitrarily large even if the noisy data is close to the true data.

2.2 Regularization Theory

In order to get a reasonable reconstruction, we have to introduce different methods that ensure a good and stable reconstructions. These methods are often defined via functions of operators.

Definition 2.10. Let $f : \mathbb{R}^+ \to \mathbb{R}$. For compact operators, we define

$$f(K)x := \sum_i f(\sigma_i) \langle\, x, u_i \,\rangle v_i \; .$$

Of course, this definition is only well-defined for functions f for which the sum converges. We can now define *regularization methods*.

Definition 2.11. A regularization of an operator $K^\dagger$ is a family of operators $(R_\alpha)_{\alpha>0}$,

$$R_\alpha : Y \to X$$

with the following properties: there exists a map $\alpha = \alpha(\delta, y^\delta)$ such that for all $y \in dom(K^\dagger)$ and all $y^\delta \in Y$ with $\|y - y^\delta\| \le \delta$,

$$\lim_{\delta\to 0} R_{\alpha(\delta,y^\delta)} y^\delta = x^\dagger$$

and

$$\lim_{\delta\to 0} \alpha(\delta, y^\delta) = 0 \, .$$

The parameter α is called regularization parameter.

In the classical setting, regularizing operators R_α are defined via filter functions F_α:

$$R_\alpha y^\delta := \sum_{i\in\mathbb{N}} \sigma_i^{-1} F_\alpha(\sigma_i) \langle\, y^\delta, v_i \,\rangle u_i \, .$$

The requirements of Definition 2.11 have some immediate consequences on the admissible filter functions. In particular, $dom(R_\alpha) = Y$ enforces $|\sigma_i^{-1} F_\alpha(\sigma_i)| \le C$ for all i, and the pointwise convergence of R_α to $K^\dagger$ requires $\lim_{\alpha\to 0} F_\alpha(t) = 1$. Well-known regularization methods are:

(i) Truncated singular value decomposition:

$$R_\alpha y^\delta := \sum_{i}^{N} \sigma_i^{-1} \langle\, y^\delta, v_i \,\rangle u_i$$

In this case, the filter function is given by

$$F_\alpha(\sigma) := \begin{cases} 1 & \text{if } \sigma \ge \alpha \\ 0 & \text{if } \sigma < \alpha \end{cases} \, .$$

(ii) Truncated Landweber iteration: For $\beta \in (0, \frac{2}{\|K\|^2})$ and $m \in \mathbb{N}$, set

$$F_{1/m}(\lambda) = 1 - (1 - \beta\lambda^2)^m \, .$$

Here, the regularization parameter $\alpha = 1/m$ only admits discrete values.

(iii) Tikhonov regularization: Here, the filter function is given by

$$F_\alpha(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha} \, .$$

The regularized solutions of Landweber's and Tikhonov's method can also be characterized as follows:

Proposition 2.12. *The regularized solution due to Landweber,* $x^\delta_{1/m}$*, is also given by the* m*-th iterate of the Landweber iteration given by*

$$x^{n+1} = x^n + K^*(y^\delta - Kx^n) \ , \quad \text{with } x^0 = 0 \quad .$$

The regularization parameter is the reciprocal of the stopping index of the iteration.

Proposition 2.13. *The regularized solution due to Tikhonov,*

$$x^\delta_\alpha := \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \alpha} \cdot \sigma_i^{-1} \langle \, y^\delta, v_i \, \rangle u_i \ ,$$

is also the unique minimizer of the Tikhonov functional

$$J_\alpha(x) = \|y^\delta - Kx\|^2 + \alpha \|x\|^2 \ , \tag{2.5}$$

which is minimized by the unique solution of the equation

$$(K^*K + \alpha I)x = K^* y^\delta \ .$$

Tikhonov's variational formulation (2.5) is important as it allows generalizations towards nonlinear operators as well as to sparse reconstructions. As mentioned above, regularization methods also require proper parameter choice rules.

Proposition 2.14. *The Tikhonov regularization combined with one of the parameter choice rules*
a) $\alpha(\delta, y^\delta) \to 0$ *and* $\frac{\delta^2}{\alpha(\delta, y^\delta)} \to 0$
b) $\alpha_*(\delta, y^\delta)$ *s.t.* $\|y^\delta - Kx^\delta_{\alpha_*}\| = \tau\delta$ *for fixed* $\tau > 1$ *(discrepancy principle)*
is a regularization method.

Proposition 2.15. *Let* $\tau > 1$*. If the Landweber iteration is stopped after* m_* *iterations, where* m_* *is the first index with*

$$\|y^\delta - Kx^{m_*}\| \le \tau\delta < \|y^\delta - Kx^{m*-1}\| \quad (discrepancy\ principle) \ ,$$

then the iteration is a regularization method with $R_{\frac{1}{m_*}} y^\delta = x^{m_*}$*.*

The last two propositions show that the regularized solutions for Tikhonov's or Landweber's method converge towards the true solution provided a proper parameter choice rule was applied. However, no result on the speed of convergence is provided. Due to Bakhushinsky one rather has

Proposition 2.16. *Let* $x^\delta_\alpha = R_\alpha y^\delta, R_\alpha$ *be a regularization method. Then the convergence of* $x^\delta_\alpha \to x^\dagger$ *can be arbitrary slow.*

To overcome this drawback, we have to assume a certain regularity of the solution. Indeed, convergence rates can be achieved provided the solution fulfills a so-called source-conditions. Here we limit ourselves to the Hölder-type source conditions,

$$x^\dagger = (K^*K)^{\frac{\nu}{2}} w, \text{ i.e. } x^\dagger \in range(K^*K)^{\frac{\nu}{2}} \subseteq dom(K) = X, \nu > 0 \; .$$

Definition 2.17. A regularization method is called order optimal if for a given parameter choice rule the estimate

$$\|x^\dagger - x^\delta_{\alpha(\delta,y^\delta)}\| = \mathcal{O}(\delta^{\frac{\nu}{\nu+1}}) \tag{2.6}$$

holds for all $x^\dagger = (K^*K)^{\frac{\nu}{2}} w$ and $\|y^\delta - y\| \leq \delta$.

It turns out that for $x^\dagger = (K^*K)^{\frac{\nu}{2}} w$ this is actually the best possible convergence rate, no method can do better. Also, we have $\delta^{\frac{\nu}{\nu+1}} > \delta$ for $\delta < 1$, i.e., we always loose some information in the reconstruction procedure.

Proposition 2.18. *Tikhonov regularization and Landweber iteration together with the discrepancy principle are order optimal.*

3 Nonlinear Approximation for Linear Ill-Posed Problems

Within this section we consider linear inverse problems and construct for them a Landweber-like algorithm for the sparse recovery of the solution x borrowing "leafs" from nonlinear approximation. The classical Landweber iteration provides in combination with suitable regularization parameter rules an order optimal regularization scheme (for the definition, see Eq. (2.6)). 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 we define an adaptive variant of Landweber's iteration from which we show regularization properties for exact and noisy data that hold in combination with an adequate refinement/stopping rule (a-priori as well as a-posteriori principles). The results presented in this Section where first published in [47]

3.1 Landweber Iteration and Its Discretization

The Landweber iteration is a gradient method for the minimization of $\|y^\delta - Ax\|^2$ and is therefore given through

$$x^{n+1} = x^n + \gamma A^*(y^\delta - Ax^n) \; . \tag{3.1}$$

As it can be retrieved, e.g. in [28], iteration (3.1) is for $0 < \gamma < 2/\|A\|^2$ a linear regularization method as long as the iteration is truncated at some finite index n_*. In order to identify the optimal truncation index n_*, one may apply either an a-priori or an a-posteriori parameter rule. The Landweber method (3.1) is an order optimal linear regularization method, see [28], if the iteration is truncated at the a-priori chosen iteration index

$$n_* = \lfloor \gamma \big(2\frac{\gamma}{\nu}e\big)^{\nu/(\nu+1)} \big(\frac{\rho}{\delta}\big)^{2/(\nu+1)} \rfloor \,, \tag{3.2}$$

where the common notation $\lfloor p \rfloor$ denotes the smallest integer less or equal p. Here, we have assumed that the solution $x^\dagger$ of our linear equation admits the smoothness condition

$$x^\dagger = (A^*A)^{\nu/(\nu+1)}v, \qquad \|v\| \le \rho \,.$$

If n_* is chosen as suggested in (3.2), then optimal convergence order with respect to $x^\dagger$ can be achieved. This proceeding, however, needs exact knowledge of the parameters $\nu,\ \rho$ in the source condition. This shortfall can be avoided when applying Morozov's discrepancy principle. This principle performs the iteration as long as

$$\|Ax^n - y^\delta\| > \tau\delta \tag{3.3}$$

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

$$\|Ax^{n_*} - y^\delta\| \le \tau\delta \tag{3.4}$$

is fulfilled for the first time. The regularization properties of this principle were investigated in [20]. The authors have shown that, as long as (3.3) 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 (3.1). This can be retracted in details in [38] where a discretization of the form

$$x^{n+1} = x^n + \gamma A^*_{r^\delta(n)}(y^\delta - A_{r^\delta(n)}x^n) \tag{3.5}$$

was suggested. The basic idea in [38] is the introduction of approximations $A_{r^\delta(m)}$ to the operator A that are updated/refined in dependence on a specific discrepancy principle.

Iteration (3.5) acts in the infinite dimensional Hilbert space X. To treat (3.5) numerically, we have to discretize the inverse problem which means that we have to find a discretized variant of (3.1) through the discretization of the corresponding normal equation of $\|y^\delta - Ax\|^2$. To this end, we assume that we have for the underlaying space X a preassigned countable system of functions $\{\phi_\lambda : \lambda \in \Lambda\} \subset X$ at our disposal for which there exist constants C_1, C_2 with $0 < C_1 \le C_2 < \infty$ such that for all $x \in X$,

$$C_1\|x\|_X^2 \le \sum_{\lambda\in\Lambda} |\langle x, \phi_\lambda\rangle|^2 \le C_2\|x\|_X^2. \tag{3.6}$$

For such a system, which is often referred to as a frame for X, see [6] for further details, we may consider the operator $\mathcal{F} : X \to \ell_2$ via $x \mapsto c = \{\langle x, \phi_\lambda \rangle\}_{\lambda\in\Lambda}$ with adjoint $\mathcal{F}^* : \ell_2 \to X$ via $c \mapsto \sum_{\lambda\in\Lambda} c_\lambda \phi_\lambda$. The operator $\mathcal{F}$ is often referred in the literature to as the analysis operator, whereas $\mathcal{F}^*$ is referred to as the synthesis operator. The composition of both, $\mathcal{F}^*\mathcal{F}$, is called the frame operator which is by condition (3.6) an invertible map; guaranteeing that each $x \in X$ can be reconstructed from its moments $\langle x, \phi_\lambda \rangle$. Moreover, there is for every $x \in X$ at least one sequence c such that $x = \mathcal{F}^* c$. Consequently, we can define

$$S = \mathcal{F}A^*A\mathcal{F}^* \ , \quad x = \mathcal{F}^* c \quad \text{and} \quad g^\delta = \mathcal{F}A^* y^\delta$$

leading to the discretized normal equation

$$Sc = g^\delta \ . \tag{3.7}$$

An approximate solution for (3.7) can then be derived by the corresponding sequence space Landweber iteration,

$$c^{n+1} = c^n + \gamma(g^\delta - Sc^n) \ . \tag{3.8}$$

Note that the operator $S : \ell_2(\Lambda) \to \ell_2(\Lambda)$ is symmetric but through the ill-posedness of A not boundedly invertible on $\ell_2(\Lambda)$ (even on the subspace Ran F). This is one major difference to [50] in which the invertibility of S on Ran F was substantially used to ensure the convergence of the Landweber iteration. Since we can neither handle the infinite dimensional vectors c^n and g^δ nor apply the infinite dimensional matrix S, iteration (3.8) is not a practical algorithm. To this end, we need to study the convergence and regularization properties of the iteration in which c^n, g^δ and S are approximated by finite length objects. Proceeding as suggested [50], we assume that we have the following three routines at our disposal:

- $\boldsymbol{RHS}_\varepsilon[y] \to g_\varepsilon$. This routine determines a finitely supported $g_\varepsilon \in \ell_2(\Lambda)$ satisfying
$$\|g_\varepsilon - \mathcal{F}A^* y\| \le \varepsilon \ .$$
- $\boldsymbol{APPLY}_\varepsilon[c] \to w_\varepsilon$. This routine determines, for a finitely supported $c \in \ell_2(\Lambda)$ and an infinite matrix S, a finitely supported w_ε satisfying
$$\|w_\varepsilon - Sc\| \le \varepsilon \ .$$
- $\boldsymbol{COARSE}_\varepsilon[c] \to c_\varepsilon$. This routine creates, for a finitely supported with $c \in \ell_2(\Lambda)$, a vector c_ε by replacing all but N coefficients of c by zeros such that
$$\|c_\varepsilon - c\| \le \varepsilon \ ,$$
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 [10, 50]. For the sake of more flexibility in our proposed approach, we allow (in contrast to classical setup suggested in [50]) ε to be different within each iteration step and sometimes different for each of the three routines. Consequently, we set $\varepsilon = \varepsilon_n^R$ for the routine $\boldsymbol{RHS}_\varepsilon[\cdot]$, $\varepsilon = \varepsilon_n^A$ for $\boldsymbol{APPLY}_\varepsilon[\cdot]$ and, finally, $\varepsilon = \varepsilon_n^C$ for $\boldsymbol{COARSE}_\varepsilon[\cdot]$. The subscript n of the created error tolerance or so-called refinement sequences $\{\varepsilon_n^C\}_{n\in\mathbb{N}}$, $\{\varepsilon_n^A\}_{n\in\mathbb{N}}$ and $\{\varepsilon_n^R\}_{n\in\mathbb{N}}$ 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 converging to zero and have to 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 $r^\delta(n) = n$ that chooses for each iteration n the preselected error tolerances ε_n^C, ε_n^A and ε_n^R. Choosing proper refinement strategies $r^\delta(n)$ enables us to establish convergence results and, thanks to the introduced subtleness, several desired regularization results.

For ease of notation we write, if not otherwise stated, instead of $\varepsilon_{r^\delta(n)}^{\{C,A,R\}}$ just the index $r^\delta(n)$, i.e. we abbreviate

$$\boldsymbol{COARSE}_{\varepsilon^C_{r^\delta(n)}}[\cdot],\ \boldsymbol{APPLY}_{\varepsilon^A_{r^\delta(n)}}[\cdot],\ \text{and}\ \boldsymbol{RHS}_{\varepsilon^R_{r^\delta(n)}}[\cdot]$$

by

$$\boldsymbol{COARSE}_{r^\delta(n)}[\cdot],\ \boldsymbol{APPLY}_{r^\delta(n)}[\cdot],\ \text{and}\ \boldsymbol{RHS}_{r^\delta(n)}[\cdot]\ .$$

Note, this does not mean the same accuracy for all three routines, it just means the same index for the accuracy/refinement sequences.

Summarizing the last steps results in the following inexact/approximative variant of (3.8)

$$\tilde{c}^{\,n+1} = \boldsymbol{COARSE}_{r^\delta(n)}\big[\tilde{c}^{\,n} - \gamma\boldsymbol{APPLY}_{r^\delta(n)}[\tilde{c}^{\,n}] + \gamma\boldsymbol{RHS}_{r^\delta(n)}[y^\delta]\big]\ . \quad (3.9)$$

3.2 Regularization Theory for A-Priori Parameter Rules

As mentioned above, the a-priori parameter rule (3.2) for the exact Landweber iteration (3.1) yields an order optimal regularization scheme. The natural question is whether the same holds true for the inexact (nonlinear and adaptive) Landweber iteration (3.9). 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 (3.1) and the inexact iteration (3.9).

Lemma 3.1. *Assume, $c^0 = \tilde{c}^{\,0}$. Then, for all $n \geq 0$,*

$$\|c^{n+1} - \tilde{c}^{\,n+1}\| \leq \gamma \sum_{i=0}^{n} (1+\gamma\|S\|)^i (\varepsilon^C_{r^\delta(n-i)}/\gamma + \varepsilon^A_{r^\delta(n-i)} + \varepsilon^R_{r^\delta(n-i)}). \tag{3.10}$$

The latter lemma allows now to prove that the truncated inexact Landweber iteration (3.9) is an order optimal regularization method. The regularization method R_α can be described with the help of an adequate refinement map r^δ and the a-priori parameter rule (3.2).

Definition 3.2 (Regularization method with a-priori parameter rule).

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

ii) for $\delta > 0$ with $\|y^\delta - y\| \leq \delta$ derive the truncation index $n_*(\delta,\rho)$ as in (3.2),

iii) define the quantities

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

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

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

v) define the regularization

$$R_\alpha g^\delta := \mathcal{F}^* \tilde{c}^{\,\delta}_{\,n_*}$$

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

Theorem 3.3 (Regularization result). *Let the truncation index $n_* = n_*(\delta,\rho)$ be as in* (3.2). *Then, the inexact Landweber iteration* (3.9) *truncated at index n_* and updated with the refinement strategy r^δ (satisfying iv) in* Definition 3.2*) yields for $\alpha(\delta,\rho) = 1/n_*(\delta,\rho)$ a regularization method R_α, which is for all $\nu > 0$ and $0 < \gamma < 2/\|S\|^2$ order optimal.*

3.3 Regularization Theory by A-Posteriori Parameter Rules

The exact Landweber iteration (3.1) combined with the discrepancy principle (3.3) and (3.4) yields a regularization method. In what follows we show how this result carries over to (3.9).

The application of the discrepancy principle (3.3) and (3.4) requires a frequent evaluation of the residual discrepancy $\|Ax^n - y^\delta\|$. Therefore, we have to propose a function that is numerical implementable and approximates the residual, preferably by means of $\boldsymbol{APPLY}$ and $\boldsymbol{RHS}$.

Definition 3.4. For some $y \in Y$, $c \in \ell_2(\Lambda)$ and some $n \geq 0$ the approximate discrepancy is defined by

$$(\boldsymbol{RES}_n[c,y])^2 := \langle \boldsymbol{APPLY}_n[c], c\rangle - 2\langle \boldsymbol{RHS}_n[y], c\rangle + \|y\|^2. \tag{3.11}$$

The following lemma gives a result on the distance between the exact function space residual discrepancy $\|Ax - y\|$ and its inexact version $\boldsymbol{RES}_n[c,y]$.

Lemma 3.5. *For $c \in \ell_2(\Lambda)$ with $Fc = x$, $y \in Y$ and some integer $n \geq 0$ it holds*

$$|\ \|Ax - y\|^2 - (\boldsymbol{RES}_n[c,y])^2\ | \leq (\varepsilon_n^A + 2\varepsilon_n^R)\,\|c\|\,. \tag{3.12}$$

To achieve convergence of (3.9), we have to elaborate under which conditions a decay of the approximation errors $\|\tilde{c}_{\ n} - c^\dagger\|$ can be ensured.

Lemma 3.6. *Let $\delta > 0$, $0 < c < 1$, $0 < \gamma < 2/(3\|S\|)$ and $n_0 \geq 1$. If there exists for $0 \leq n \leq n_0$ a refinement strategy $r^\delta(n)$ such that $\boldsymbol{RES}_{r^\delta(n)}[\tilde{c}^{\ n}, y^\delta]$ fulfills*

$$c(\boldsymbol{RES}_{r^\delta(n)}[\tilde{c}^{\ n}, y^\delta])^2 > \frac{\delta^2 + C_{r^\delta(n)}(\tilde{c}^{\ n})}{1 - \frac{3}{2}\gamma\|S\|}\ , \tag{3.13}$$

then, for $0 \leq n \leq n_0$, the approximation errors $\|\tilde{c}^{\ n} - c^\dagger\|$ decrease monotonically.

The above Lemma 3.6 holds in particular for exact data, i.e. $\delta = 0$. In this case, condition (3.13) simplifies to

$$c(\boldsymbol{RES}_{r(m)}[\tilde{c}_{\ m}, y])^2 \geq \frac{C_{r(n)}(\tilde{c}^{\ n})}{1 - \frac{3}{2}\gamma\|S\|}\ . \tag{3.14}$$

To prove convergence, we follow the suggested proceeding in [38] and introduce an updating rule (U) for the refinement strategy r:

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

$$c(\boldsymbol{RES}_{r(0)}[\tilde{c}_0, y])^2 \geq \frac{C_{r(0)}(\tilde{c}_0)}{1 - \frac{3}{2}\gamma\|S\|}, \tag{3.15}$$

if $r(0)$ with (3.15) does not exist, stop the iteration, set $n_0 = 0$.

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

$$c(\boldsymbol{RES}_{r(n-1)}[\tilde{c}^{\,n}, y])^2 \geq \frac{C_{r(n-1)}(\tilde{c}^{\,n})}{1 - \frac{3}{2}\gamma\|S\|}, \tag{3.16}$$

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

U(iii) if

$$c(\boldsymbol{RES}_{r(n-1)}[\tilde{c}^{\,n}, y])^2 < \frac{C_{r(n-1)}(\tilde{c}^{\,n})}{1 - \frac{3}{2}\gamma\|S\|}, \tag{3.17}$$

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

$$c(\boldsymbol{RES}_{r(n-1)+j}[\tilde{c}^{\,n}, y])^2 \geq \frac{C_{r(n-1)+j}(\tilde{c}^{\,n})}{1 - \frac{3}{2}\gamma\|S\|}, \tag{3.18}$$

U(iv) if there is no integer j with (3.18), then stop the iteration, set $n_0 = n$.

Lemma 3.7. *Let $\delta = 0$ and $\{\tilde{c}^{\,n}\}_{n\in\mathbb{N}}$ be the sequence of iterates* (3.9)*. Assume the updating rule* (U) *for r was applied. Then, if the iteration never stops,*

$$\sum_{n=0}^{\infty}(\boldsymbol{RES}_{r(n)}[\tilde{c}^{\,n}, y])^2 \leq \frac{1}{\beta(1-c)\big(1 - \frac{3}{2}\gamma\|S\|\big)}\,\|\tilde{c}^{\,n} - c^\dagger\|^2. \tag{3.19}$$

If the iteration stops after n_0 steps,

$$\sum_{n=0}^{n_0-1}(\boldsymbol{RES}_{r(n)}[\tilde{c}^{\,n}, y])^2 \leq \frac{1}{\beta(1-c)\big(1 - \frac{3}{2}\gamma\|S\|\big)}\,\|\tilde{c}^{\,0} - c^\dagger\|^2. \tag{3.20}$$

Combining the monotone decay of the approximation errors and the uniform boundedness of the accumulated discrepancies enables strong convergence of iteration (3.9) towards a solution of the inverse problem for exact data $y^\delta = y$.

Theorem 3.8. *Let $x^\dagger$ denote the generalized solution of the given inverse problem. Suppose $\tilde{c}^{\,n}$ is computed* (3.9) *with exact data y in combination with updating rule*

(U) *for the refinement strategy* r. *Then, for arbitrarily chosen* $\tilde{c}^{\,0}$ *the sequence* $\tilde{c}^{\,n}$ *converges in norm, i.e.*

$$\lim_{n\to\infty} \tilde{c}^{\,n} = c^\dagger$$

with

$$x^\dagger = \mathcal{F}^* c^\dagger\ .$$

The convergence of the inexact Landweber iteration for noisy data relies on a comparison between the individual noise free and noisy iterations. For a comparison it is essential to analyze the δ-dependence of $\boldsymbol{COARSE}$, $\boldsymbol{APPLY}$ and $\boldsymbol{RHS}$; in particular for $\delta \to 0$. For a given error level $\|v^\delta - v\| \leq \delta$ the routines (here just exemplarily stated for $\boldsymbol{COARSE}$, but must hold for all three routines) should fulfill for any fixed $\varepsilon > 0$

$$\|\boldsymbol{COARSE}_\varepsilon(v^\delta) - \boldsymbol{COARSE}_\varepsilon(v)\| \to 0 \quad \text{as} \quad \delta \to 0. \tag{3.21}$$

To ensure (3.21), the three routines as proposed in [50] must be slightly adjusted, which is demonstrated in great details for the $\boldsymbol{COARSE}$ routine only but must be done for $\boldsymbol{APPLY}$ and $\boldsymbol{RHS}$ accordingly.

The definition of $\boldsymbol{COARSE}$ as proposed in [50] (with a slight modified ordering of the output entries) is as follows

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

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

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

$$\boldsymbol{V}_k := \{v_i \in \boldsymbol{V} : 2^{-k-1}\|v\| < |v_i| \leq 2^{-k}\|v\|\}, \tag{3.22}$$

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 v. Denote the vector obtained by subsequently extracting the elements of $\boldsymbol{V}_0, \ldots, \boldsymbol{V}_q$ by $\gamma(v)$.

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

$$\|v - v_\varepsilon\|^2 = \sum_{i>l} |\gamma_i(v)|^2 < \varepsilon^2. \tag{3.23}$$

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

Keeping the original order of the coefficients of v in $\boldsymbol{V}_k$, the output vector of $\boldsymbol{COARSE}$ becomes unique. This “natural” ordering does not cause any extra computational cost.

The goal is to construct a noise dependent output vector that converges to the noise free output vector as $\delta \to 0$. To achieve this, the uniqueness of $\boldsymbol{COARSE}$ must be ensured. The non-uniqueness of $\boldsymbol{COARSE}$ is through the bin sorting procedure which is naturally non-unique as long as the input vectors are noisy (i.e. different noisy versions of the same vector result in significantly different output vectors). This leads to the problem that the index in $\gamma(v^\delta)$ of some noisy element v_i^δ can differ to the index in $\gamma(v)$ of its noise free version v_i. To overcome this drawback for (at least) sufficiently small δ, we define a noise dependent version $\boldsymbol{COARSE}^\delta$.

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

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

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

$$\boldsymbol{V}_k^\delta := \{v_i^\delta \in \boldsymbol{V}^\delta : 2^{-k-1}(\|v^\delta\| + \delta) + \delta < |v_i^\delta| \le 2^{-k}(\|v^\delta\| + \delta) + \delta\}, \tag{3.24}$$

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 v^δ. Denote the vector obtained by the bin sorting process by $\gamma^\delta(v^\delta)$.

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

$$\|v^\delta - v_\varepsilon^\delta\|^2 = \|v^\delta\|^2 - \sum_{1 \le i \le l^\delta} |\gamma_i^\delta(v^\delta)|^2 < \varepsilon^2 - (l^\delta + 1)\delta(2\|v^\delta\| + \delta). \tag{3.25}$$

The latter definition of $\boldsymbol{COARSE}^\delta$ enables us to achieve the desired property (3.21).

Lemma 3.9. *Given $\varepsilon > 0$ and $\delta > 0$. For arbitrary finite length vectors v, $v^\delta \in \ell_2$ with $\|v^\delta - v\| \le \delta$, the routine $\boldsymbol{COARSE}^\delta$ is convergent in the sense that*

$$\|\boldsymbol{COARSE}_\varepsilon^\delta[v^\delta] - \boldsymbol{COARSE}_\varepsilon[v]\| \to 0 \quad \text{as} \quad \delta \to 0. \tag{3.26}$$

Achieving convergence of the inexact iteration, we introduce as for the noise free situation an updating rule which we denote (D). The updating rule (D) is based on the refinement strategy $r^\delta(n)$.

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

$$c(\boldsymbol{RES}_{r^\delta(0)}[\tilde{c}\,_0^\delta, y^\delta])^2 \geq \frac{\delta^2 + C_{r^\delta(0)}(\tilde{c}\,_0^\delta)}{1 - \frac{3}{2}\beta\|S\|}, \tag{3.27}$$

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

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

$$c(\boldsymbol{RES}_{r^\delta(n-1)}[\tilde{c}\,_n^\delta, y^\delta])^2 \geq \frac{\delta^2 + C_{r^\delta(n-1)}(\tilde{c}\,_n^\delta)}{1 - \frac{3}{2}\beta\|S\|}, \tag{3.28}$$

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

D(iii) if

$$c(\boldsymbol{RES}_{r^\delta(n-1)}[\tilde{c}\,_n^\delta, y^\delta])^2 < \frac{\delta^2 + C_{r^\delta(n-1)}(\tilde{c}\,_n^\delta)}{1 - \frac{3}{2}\beta\|S\|}, \tag{3.29}$$

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

$$c(\boldsymbol{RES}_{r^\delta(n-1)+j}[\tilde{c}\,_n^\delta, y^\delta])^2 \geq \frac{\delta^2 + C_{r^\delta(n-1)+j}(\tilde{c}\,_n^\delta)}{1 - \frac{3}{2}\beta\|S\|} \tag{3.30}$$

and

$$C_{r^\delta(m-1)+j}(\tilde{c}\,_m^\delta) > c_1\delta^2. \tag{3.31}$$

D(iv) if (3.29) holds and no j with (3.30),(3.31) exists, then stop the iteration, set $n_*^\delta = n$.

Theorem 3.10. *Let $x^\dagger$ be the solution of the inverse problem for exact data $y \in \mathrm{Ran}\, A$. Suppose that for any $\delta > 0$ and y^δ with $\|y^\delta - y\| \leq \delta$ the adaptive approximation $\tilde{c}\,_n^\delta$ is derived by the inexact Landweber iteration* (3.9) *in combination with rule* (D) *for r^δ and stopping index n_*^δ. Then, the family of R_α defined through*

$$R_\alpha y^\delta := \mathcal{F}^* \tilde{c}\,_{n_*^\delta}^\delta \quad \text{with} \quad \alpha = \alpha(\delta, y^\delta) = \frac{1}{n_*^\delta}$$

yields a regularization of the ill-posed operator A, i.e. $\|R_\alpha y^\delta - x^\dagger\|_X \to 0$ as $\delta \to 0$.

4 Tikhonov Regularization with Sparsity Constraints

In this section we turn now to nonlinear inverse and ill-posed problems. The focus is on the generalization of Tikhonov regularization as it was introduced in Section 2 (see formula (2.5)) to nonlinear problems. In particular, we consider those operator equations in which the solution x has a *sparse* series expansion $x = \sum_{\lambda\in\Lambda} c_\lambda \phi_\lambda$ with respect to a preassigned basis or frame, i.e. the series expansion of x has only a very small number of non-vanishing coefficients c_λ, or that x is compressible (meaning that x can be well-approximated by a sparse series expansion).

4.1 Regularization Result for A-Priori Parameter Rules

We consider the operator equation $F(x) = y$ and assume F is possibly ill-posed and maps between Hilbert spaces X and Y and we suppose there are only noisy data y^δ with $\|y^\delta - y\| \leq \delta$ available. The natural generalization of Tikhonov's variational formulation is then given by

$$J_\alpha(x) = \|F(x) - y^\delta\|^2 + \alpha\|x\|^2 . \tag{4.1}$$

The second term determines the properties of the solution. In the given setting the penalty term is a quadratic Hilbert space norm ensuring finite energy of the solution. The minimizer is due to the convexity and differentiability of $\|\cdot\|^2$ also very easy to compute. However, for certain classes of inverse problems, e.g. in medical or astrophysical imaging or signal peak analysis, such Hilbert space constraints seem not to be best suited, because they lead to over-smoothed solutions implying that jumps and edges cannot be nicely reconstructed. Therefore, alternatives are required that can perform much better. An alternative that may circumvent the mentioned drawbacks are so-called sparsity measures. Prominent examples of sparsity measures are ℓ_p-norms, $0 < p < 2$, on the coefficients of the series expansions of the solution to be reconstructed. But also much more general constraints such as the wide class of convex, one–homogeneous and weakly lower semi-continuous constraints are possible, see e.g. [3, 33, 34, 36, 49] or [9, 11].

In what follows we restrict ourselves to ℓ_p-norm constraints. Once a frame is preassigned, we know that for every $x \in X$ there is a sequence c such that $x = \mathcal{F}^*c$, and therefore the given operator equation can be expressed as $F(\mathcal{F}^*c) = y$. Consequently, we can define, for a given a-priori guess $\bar{c} \in \ell_2(\Lambda)$, an adequate Tikhonov functional by

$$J_\alpha(c) = \|F(\mathcal{F}^*c) - y^\delta\|^2 + \alpha\Psi(c, \bar{c}) \tag{4.2}$$

with minimizer

$$c_\alpha^\delta := \arg\min_{c\in\ell_2(\Lambda)} J_\alpha(c) .$$

To specify Ψ, we define

$$\Psi_{p,w}(c) := \left(\sum_{\lambda\in\Lambda} w_\lambda |c_\lambda|^p\right)^{1/p} ,$$

where $w = \{w_\lambda\}_{\lambda\in\Lambda}$ is a sequence of weights with $0 < C < w_\lambda$. A popular choice for the ansatz system $\{\phi_\lambda : \lambda \in \Lambda\}$ are wavelet bases or frames. In particular, for orthonormal wavelet bases and for properly chosen weights w one has $\Psi_{p,w}(c) \sim \|c\|_{B^s_{p,p}}$, where $B^s_{p,p}$ denotes a standard Besov space. In this section we restrict the analysis to either $\Psi(c,\bar{c}) = \Psi_{p,w}(c-\bar{c})$ or $\Psi(c,\bar{c}) = \Psi^p_{p,w}(c-\bar{c})$.

For c^δ_α as a minimizer of (4.2) we can achieve for any of the introduced sparsity measures regularization properties if the following assumptions hold true:

(i) F is strongly continuous, i.e. $c^n \overset{w}{\to} c \Rightarrow F(c^n) \to F(c)$,

(ii) a-priori parameter available with $\alpha(\delta) \to 0$ and $\delta^2/\alpha(\delta) \to 0$ as $\delta \to 0$,

(iii) $x^\dagger$ as well $\bar{c}$ have finite value of $\Psi_{p,\omega}$.

Here, we denote by the symbol $\overset{w}{\to}$ the weak convergence.

Theorem 4.1. *Suppose (i) and (iii) hold and that we are given a sequences $\delta_k \to 0$ and $\alpha(\delta_k)$ with (ii). Then the sequence of minimizers $c^{\delta_k}_{\alpha_k}$ has a convergent subsequence that converges with respect to Ψ towards a solution of $F(\mathcal{F}^*c) = y$. If the solution is unique, the whole sequence converges with respect to Ψ, i.e.*

$$\lim_{k\to\infty} \Psi(c^{\delta_k}_{\alpha_k} - c^\dagger) = 0\,. \tag{4.3}$$

Consequently, the variational Tikhonov approach with properly chosen sparsity constraints is a regularization method.

4.2 Convergence Rates for A-Priori Parameter Rules

The convergence in (4.3) can be arbitrarily slow. Therefore, conditions for establishing convergence rates need to be achieved. As the analysis that is required for nonlinear operator equations can be under several conditions on the operator F reduced to the study of the linear operator case, we limit the discussion to the linear case for which convergence rates can be shown. The results within this Section have been first published in [48].

Consider the linear and ill-posed operator equation

$$\begin{aligned} \tilde{A}x &= g \\ \tilde{A} &: X_{p,\omega} \to L_2(\Omega)\,. \end{aligned} \tag{4.4}$$

Here, $X_{p,w}$ denotes a Banach space which is a subspace of $L_2(\Omega)$, with parameters $p \in (1,2)$ and $w = \{w_\lambda\}_{\lambda\in\Lambda}$, where Ω is a bounded open subset of $\mathbb{R}^d$, with $d \geq 1$, and Λ is an index set of (possibly tuples of) integer indices. Although one could employ more general separable Hilbert spaces than $L_2(\Omega)$, we consider here the Lebesgue space case, for simplicity.

We are in particular interested in the reconstruction of solutions of (4.4) that admit a sparse structure with respect to a given basis in the Banach space $X_{p,w}$. In these cases it is desirable to choose a regularization method that also promotes a sparse reconstruction. For instance, suitable choices for the spaces $X_{p,w}$ are the Besov spaces $B^s_{p,p}$ with $p \in (1,2)$, in case of a sufficiently smooth wavelet basis and properly chosen weights - see, e.g., [5], [37] for detailed discussions.

Instead of solving the above equation in a function space setting, we will transform it into a sequential setting. By choosing a suitable orthonormal basis $\Phi = \{\phi_\lambda : \lambda \in \Lambda\}$ for the space $L_2(\Omega)$, both x and $\tilde{A}x$ can be expressed with respect to Φ. Thus,

$$\tilde{A}x = \sum_{\lambda'} \sum_{\lambda} \langle x, \phi_\lambda \rangle \langle \tilde{A}\phi_\lambda, \phi_{\lambda'} \rangle \, \phi_{\lambda'} \, . \tag{4.5}$$

Defining the infinite dimensional matrix A and vectors c, y by

$$A = (\langle \tilde{A}\phi_\lambda, \phi_{\lambda'} \rangle)_{\lambda,\lambda' \in \Lambda}, \; c = (\langle c, \phi_\lambda \rangle)_{\lambda \in \Lambda}, \; y = (\langle g, \phi_\lambda \rangle)_{\lambda \in \Lambda} \, , \tag{4.6}$$

equation (4.4) can be reformulated as an (infinite) linear system

$$Ac = b. \tag{4.7}$$

To specify the spaces $X_{p,w}$, we define for a given orthonormal basis Φ and positive weights w

$$x \in X_{p,w} \Longleftrightarrow \sum_\lambda w_\lambda |\langle x, \phi_\lambda \rangle|^p < \infty \, ,$$

i.e. c belongs to the weighted sequence space $\ell_{p,w}$, where

$$\ell_{p,w} = \left\{ c = \{c_\lambda\}_{\lambda \in \Lambda} : \|c\|_{p,w} = \left(\sum_\lambda w_\lambda |c_\lambda|^p \right)^{\frac{1}{p}} < \infty \right\}.$$

Since $\ell_p \subset \ell_q$ with $\|c\|_q \le \|c\|_p$ for $p \le q$, one also has $\ell_{p,w} \subset \ell_{q,w'}$ for $p \le q$ and $w' \le w$. In particular, if the sequence of weights is positive and bounded from below, i.e., $0 < C \le w_\lambda$ for some $C > 0$, then $\ell_{p,w} \subset \ell_2$ for $p \le 2$.

With the above discretization, we consider the sequence space operator equation

$$\begin{aligned} Ac &= y \\ A &: \ell_{p,w} \to \ell_2, \end{aligned} \tag{4.8}$$

where A is a linear and bounded operator. Now we are prepared to investigate convergence rates for Tikhonov regularization with sparsity constraints, where the approximation of the solution is obtained as a minimizer of

$$J_\alpha(c) = \|Ac - y^\delta\|^2 + 2\alpha \Psi^p_{p,w}(c) \, , \tag{4.9}$$

with regularization parameter $\alpha > 0$. Note that the function $\Psi^p_{p,w}$ is strictly convex since the p-powers of the norms are so. In addition, the function $\Psi^p_{p,w}$ is Fréchet differentiable. In order to obtain convergence rates we need the following source conditions,

(SC) ${\Psi^p_{p,w}}'(c^\dagger) = A^*v$, for some $v \in \ell_2$.

(SC I) ${\Psi^p_{p,w}}'(c^\dagger) = A^*A\hat{v}$, for some $\hat{v} \in \ell_{p,w}$.

For the above given source conditions we get the following convergence rates:

Proposition 4.2. *Assume that the noisy data y^δ fulfill $\|y - y^\delta\| \le \delta$ and that $p \in (1,2)$.*
i) If (SC) *and $\alpha \sim \delta$, then the following error estimates hold for the minimizer c^δ_α of* (4.9)*:*

$$\|c^\delta_\alpha - c^\dagger\|_{p,w} = \mathcal{O}(\delta^{\frac{1}{2}}), \quad \|Ac^\delta_\alpha - y\| = \mathcal{O}(\delta).$$

ii) If (SC I) *and $\alpha \sim \delta^{\frac{2}{p+1}}$, then*

$$\|c^\delta_\alpha - c^\dagger\|_{p,w} = \mathcal{O}(\delta^{\frac{p}{p+1}}), \quad \|Ac^\delta_\alpha - y\| = \mathcal{O}(\delta).$$

Recently it is shown in [26] that under the assumption that $c^\dagger$ is sparse and (SC) holds, the convergence rate is $\mathcal{O}(\delta^{\frac{1}{p}})$ for $p \in [1,2)$ (thus, up to $\mathcal{O}(\delta)$) with respect to the ℓ_2 norm of $c^\delta_\alpha - c^\dagger$ (which is weaker than the $\ell_{p,w}$ norm for $p < 2$). These rates are already higher, when $p < 1.5$, than the "superior limit" of $\mathcal{O}(\delta^{\frac{2}{3}})$ established for quadratic regularization. This indicates that the assumption of sparsity is a very strong source condition. Next we give a converse results for the first source condition, which shows that the above given convergence rate can only hold if the source condition is fulfilled.

Proposition 4.3. *If $\|y - y^\delta\| \le \delta$, the rate $\|Ac^\delta_\alpha - y\| = \mathcal{O}(\delta)$ holds and c^δ_α converges to $c^\dagger$ in the $\ell_{p,w}$ weak topology as $\delta \to 0$ and $\alpha \sim \delta$, then ${\Psi^p_{p,w}}'(c^\dagger)$ belongs to the range of the adjoint operator A^*.*

In what follows, we characterize sequences that fulfill the source condition (SC I). To this end we introduce the power of a sequence by

$$w^t = \{w^t_\lambda\}_{\lambda\in\Lambda}, \quad t \in \mathbb{R}\,,$$

and will consider the operator

$$A : \ell_{p',w'} \to \ell_2\,. \tag{4.10}$$

Please note that $\|\cdot\|_{p,w}$ is still used as penalty and that p, p' and w, w' are allowed to be different, respectively. In the sequel, the dual exponents to the given p, p' will be denoted by q, q'. Consider first the case $p, p' > 1$.

Proposition 4.4. *Let $p, p' > 1$, the operator A and $\Psi^p_{p,w}$ be given as above, and assume that $p \le p'$, $w' \le w$ holds true. Then a solution $c^\dagger$ of $Ac = y$ fulfilling $A^*v = {\Psi^p_{p,w}}'(c^\dagger)$ satisfies*

$$c^\dagger \in \ell_{(p-1)q',(w')^{-q'/p'}\cdot w^{q'}}\,. \tag{4.11}$$

The previous result states only a necessary condition. In order to characterize the smoothness condition in terms of spaces of sequences, we relate the spaces to $range(A^*)$:

Proposition 4.5. *Let* $p, p' > 1$*, the operator* A *and* $\Psi^p_{p,w}$ *be given as above, and assume that* $p \le p',\ w' \le w$ *holds true. Moreover, assume*

$$range(A^*) = \ell_{\tilde{q},\tilde{w}^{-\tilde{q}/\tilde{p}}} \subset \ell_{q',w'^{-q'/p'}}$$

for some $\tilde{p},\ \tilde{q} > 1$*. Then each sequence*

$$c^\dagger \in \ell_{(p-1)\tilde{q},\tilde{w}^{-\tilde{q}/\tilde{p}}\cdot w^{\tilde{q}}} \tag{4.12}$$

fulfills the smoothness condition (SC).

The above derived conditions on sequences fulfilling a source condition (SC) mean in principle that the sequence itself has to converge to zero fast enough. They can also be interpreted in terms of smoothness of an associated function: If the function system Φ in (4.5), (4.6) is formed by a wavelet basis, then the norm of a function in the Besov space $B^s_{p,p}$ coincides with a weighted ℓ_p norm of its wavelet coefficients and properly chosen weights [8]. In this sense, the source condition requires the solution to belong to a certain Besov space. The assumption on $range(A^*)$ in Proposition 4.5 then means that the range of the dual operator equals a Besov space. Similar assumptions were used for the analysis of convergence rates for Tikhonov regularization in Hilbert scales, see [31, 30, 27].

4.3 Regularization Result for A-Posteriori Parameter Rules

We deal with Morozov's discrepancy principle as an a-posteriori parameter choice rule for Tikhonov regularization with general convex penalty terms Ψ. The results presented in this Section were first published in [1]. In this framework it can be shown that a regularization parameter α fulfilling the discprepancy principle exists, whenever the operator F satisfies some basic conditions, and that for suitable penalty terms the regularized solutions converge to the true solution in the topology induced by Ψ. It is illustrated that for this parameter choice rule it holds $\alpha \to 0$, $\delta^q/\alpha \to 0$ as the noise level δ goes to 0.

We assume the operator $F : dom(F) \subset X \to Y$ between reflexive Banach spaces X, Y, with $0 \in dom(F)$, to be weakly continuous, $q > 0$ to be fixed, and that the penalty term $\Psi(x)$ fulfills the following condition.

Condition 4.6. Let $\Psi : D(\Psi) \subset X \to \mathbb{R}^+$, with $0 \in dom(\Psi)$, be a convex functional such that

(i) $\Psi(x) = 0$ if and only if $x = 0$,

(ii) Ψ is weakly lower semicontinuous (w.r.t. the Banach space topology on X),

(iii) Ψ is weakly coercive, i.e. $\|x_n\| \to \infty \Rightarrow \Psi(x_n) \to \infty$.

We want to recover solutions $x \in X$ of $F(x) = y$, where we are given y^δ with $\|y^\delta - y\| \le \delta$.

Definition 4.7. As before, our regularized solutions will be the minimizers x_α^δ of the Tikhonov-type variational functionals

$$J_\alpha(x) = \begin{cases} \|F(x) - y^\delta\|^q + \alpha\Psi(x) & \text{if } x \in dom(\Psi) \cap dom(F) \\ +\infty & \text{otherwise.} \end{cases} \tag{4.13}$$

For fixed y^δ, we denote the set of all minimizers by M_α, i.e.

$$M_\alpha = \{x_\alpha^\delta \in X \ : \ J_\alpha(x_\alpha^\delta) \le J_\alpha(x),\ \forall x \in X\} \tag{4.14}$$

We call a solution $x^\dagger$ of equation $F(x) = y$ an Ψ-minimizing solution if

$$\Psi(x^\dagger) = \min\ \{\Psi(x) : F(x) = y\},$$

and denote the set of all Ψ-minimizing solutions by $\mathcal{L}$. Throughout this paper we assume that $\mathcal{L} \neq \emptyset$.

Morozov's discrepancy principle goes now as follows.

Definition 4.8. For $1 < \tau_1 \le \tau_2$ we choose $\alpha = \alpha(\delta, y^\delta) > 0$ such that

$$\tau_1\delta \le \|F(x_\alpha^\delta) - y^\delta\| \le \tau_2\delta \tag{4.15}$$

holds for some $x_\alpha^\delta \in M_\alpha$.

Condition 4.9. Assume that y^δ satisfies

$$\|y - y^\delta\| \le \delta < \tau_2\delta < \|F(0) - y^\delta\|, \tag{4.16}$$

and that there is no $\alpha > 0$ with minimizers $x_1, x_2 \in M_\alpha$ such that

$$\|F(x_1) - y^\delta\| < \tau_1\delta \le \tau_2\delta < \|F(x_2) - y^\delta\|.$$

Then we have the following

Theorem 4.10. *If Condition 4.9 is fulfilled, then there are $\alpha = \alpha(\delta, y^\delta) > 0$ and $x_\alpha^\delta \in M_{\alpha(\delta,y^\delta)}$ such that* (4.15) *holds.*

Based on this existence result, we are able to establish regularization properties.

Condition 4.11. Let $(x^n)_{n\in\mathbb{N}} \subset X$ be such that $x^n \xrightarrow{w} \bar{x} \in X$ and $\Psi(x^n) \to \Psi(\bar{x}) < \infty$, then x^n converges to $\bar{x}$ with respect to Ψ, i.e.,

$$\Psi(x^n - \bar{x}) \to 0.$$

Remark 4.12. Choosing weighted ℓ_p-norms of the coefficients with respect to some frame $\{\phi_\lambda : \lambda \in \Lambda\} \subset X$ as the penalty term, i.e.

$$\Psi_{p,w}(x) = \|x\|_{w,p} = \Big(\sum_{\lambda\in\Lambda} w_\lambda |\langle x, \phi_\lambda\rangle|^p\Big)^{1/p}, \qquad 1 \le p \le 2, \tag{4.17}$$

where $0 < C \le w_\lambda$, satisfies Condition 4.11. Therefore the same automatically holds for $\Psi^p_{p,w}(x)$. Note that these choices also fulfill all the assumptions in Condition 4.6.

Theorem 4.13. *Let $\delta_n \to 0$ and F, Ψ satisfy the Conditions 4.6, 4.11. Assume that y^{δ_n} fulfills Condition 4.9 and choose $\alpha_n = \alpha(\delta_n, y^{\delta_n})$, $x_n \in M_{\alpha_n}$ such that* (4.15) *holds, then each sequence x_n has a subsequence that converges to an element of $\mathcal{L}$ with respect to Ψ.*

Remark 4.14. If instead of Condition 4.11 the penalty term $\Psi(x)$ satisfies the Kadec property, i.e., $x_n \xrightarrow{w} \bar{x} \in X$ and $\Psi(x_n) \to \Psi(\bar{x}) < \infty$ imply $\|x_n - \bar{x}\| \to 0$, then the convergence in Corollary 4.13 holds with respect to the norm.

Condition 4.15. For all $x^\dagger \in \mathcal{L}$ (see Definition 4.7) we assume that

$$\lim_{t\to 0^+}\inf \frac{\|F((1-t)x^\dagger) - y)^q\|}{t} = 0. \tag{4.18}$$

The following Lemma provides more insight as to the nature of Condition 4.15.

Lemma 4.16. *Let X be a Hilbert space and $q > 1$. If $F(x)$ is differentiable in the directions $x^\dagger \in \mathcal{L}$ and the derivatives are bounded in a neigbourhood of $x^\dagger$, then Condition 4.15 is satisfied.*

Theorem 4.17. *Let F, Ψ satisfy the Conditions 4.6, 4.15. Moreover, assume that data y^δ, $\delta \in (0, \delta^*)$, are given such that Condition 4.9 holds, where $\delta^* > 0$ is an arbitrary upper bound. Then the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ obtained from Morozov's discrepancy principle (see Definition 4.8) satisfies*

$$\alpha(\delta, y^\delta) \to 0 \qquad \text{and} \qquad \frac{\delta^q}{\alpha(\delta, y^\delta)} \to 0 \qquad \text{as } \delta \to 0.$$

Remark 4.18. In the proof of Theorem 4.17 we have used that $\|F(0)-y\| > 0$, which is an immediate consequence of (4.16). On the other hand, whenever $\|F(0) - y\| > 0$ we can choose

$$0 < \delta^* \leq \frac{1}{\tau_2 + 1}\|F(0) - y\|$$

and for all $0 < \delta < \delta^*$ and y^δ satisfying $\|y - y^\delta\| \leq \delta$ we obtain

$$\|F(0) - y^\delta\| \geq \|F(0) - y\| - \|y - y^\delta\| \geq \|F(0) - y\| - \delta > \tau_2\delta,$$

which is (4.16). Therefore (4.16) can be fulfilled for all δ smaller than some $\delta^* > 0$, whenever $y \neq F(0)$.

4.4 Convergence Rates for A-Posteriori Parameter Rules

Finally, we establish convergence rates with respect to the generalized Bregman distance.

Definition 4.19. Let $\partial\Psi(x)$ denote the subgradient of Ψ at $x \in X$. The generalized Bregman distance with respect to Ψ of two elements $x, z \in X$ is defined as

$$D_\Psi(x, z) = \{D_\Psi^\xi(x, z) : \xi \in \partial\Psi(z) \neq \emptyset\},$$

where

$$D_\Psi^\xi(x, z) = \Psi(x) - \Psi(z) - \langle \xi, x - z\rangle.$$

Condition 4.20. Let $x^\dagger$ be an arbitrary but fixed Ψ-minimizing solution of $F(x) = y$. Assume that the operator $F : X \to Y$ is Gâteaux differentiable and that there is $w \in Y^*$ such that

$$F'(x^\dagger)^* w \in \partial\Psi(x^\dagger). \tag{4.19}$$

Throughout the remainder of this section let $w \in Y^*$ be arbitrary but fixed fulfilling (4.19) and $\xi \in \partial\Psi(x^\dagger)$ be defined as

$$\xi = F'(x^\dagger)^* w. \tag{4.20}$$

Moreover, assume that one of the two following non-linearity conditions holds:

(i) There is $c > 0$ such that for all $x, z \in X$ it holds that

$$\langle w, F(x) - F(z) - F'(z)(x - z)\rangle \leq c\|w\|_{Y^*}\|F(x) - F(z)\|. \tag{4.21}$$

(ii) There are $\rho > 0, c > 0$ such that for all $x \in dom(F) \cap \mathcal{B}_\rho(x^\dagger)$,

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\| \leq c\, D_\Psi^\xi(x, x^\dagger), \tag{4.22}$$

and it holds that

$$c\|w\|_{Y^*} < 1. \tag{4.23}$$

Here, $\mathcal{B}_\rho(x^\dagger)$ denotes a ball around $x^\dagger$ with radius ρ.

Theorem 4.21. *Let the operator F and the penalty term Ψ be such that Condtions 4.6 and 4.20 hold. For all $0 < \delta < \delta^*$ assume that the data y^δ fulfill Condition 4.9, and choose $\alpha = \alpha(\delta, y^\delta)$ according to the discrepancy principle in Definition 4.8. Then*

$$\|F(x_\alpha^\delta) - F(x^\dagger)\| = \mathcal{O}(\delta), \qquad D_\Psi^\xi(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta). \tag{4.24}$$

5 Iterated Shrinkage for Nonlinear Ill-Posed Problems

This section is devoted to the elaboration of a numerical scheme to derive a minimizer of

$$J_\alpha(c) = \|F(\mathcal{F}^*c) - y^\delta\|^2 + \alpha\Psi(Bc)\ , \tag{5.1}$$

for some given $\alpha > 0$ and where we have assumed (for simplicity) that $\bar{c} = 0$ motivating the shorthand notation to $\Psi(c)$ for $\Psi(c, 0)$ and where B is an isometric mapping. But, in contrast to Section 4, in which the choice of Ψ was restricted to weighted ℓ_p-norms, we allow in this section a much broader range constraints, namely the wide range of positive, one–homogeneous, lower semi–continuous and convex penalty constraints, where the ℓ_p norm is just one famous example. Further important cases such as the TV measure can be found in [3, 33, 34, 36, 49]. Since we focus here on constraints that work on the basis of frame coefficients, TV–like constraints are not directly applicable here. But there is a remarkable relation between TV penalties and frame coefficient-oriented constraints which can be explained by the inclusion $B^1_{1,1} \subset BV \subset B^1_{1,1} - weak$ (in two dimensions), see for further Harmonic analysis on BV [9, 11]. This relation yields a wavelet-based near BV reconstruction when limiting to Haar frames and using a $B^1_{1,1}$ constraint, see for further elaboration [16, 17].

One additional important condition that is necessary for our further analysis is

$$\|c\|_{\ell_2} \leq \Psi(Bc). \tag{5.2}$$

To derive a minimizer of (5.1), we follow the strategies developed for nonlinear problems with quadratic penalties suggested in [44]. These concepts seem to be also adequate when dealing with sparsity, or more general, with one–homogeneous constraints. The idea goes as follows: we replace (5.1) by a sequence of functionals from which we hope that they are easier to treat and that the sequence of minimizers converge in some sense to, at least, a critical point of (5.1). To be more concrete, for some auxiliary $a \in \ell_2$, we introduce the a surrogate functional

$$J_\alpha^s(c, a) := J_\alpha(c) + C\|c - a\|^2_{\ell_2} - \|F(\mathcal{F}^*c) - F(\mathcal{F}^*a)\|^2_Y \tag{5.3}$$

and create an iteration process by:

(i) Pick c^0 and some proper constant $C > 0$

(ii) Derive a sequence $(c^n)_{n=0,1,...}$ by the iteration:

$$c^{n+1} = \arg\min_c J^s_\alpha(c, c^n) \qquad n = 0, 1, 2, \ldots \tag{5.4}$$

As a minor but relatively common restriction, convergence of iterations (5.4) can only be established when the class of operators F is restricted to (twice)Frechét differentiable operators fulfilling

$$c^n \stackrel{w}{\to} c^\star \Longrightarrow F(\mathcal{F}^* c^n) \to F(\mathcal{F}^* c^\star) \ , \tag{5.5}$$

$$F'(\mathcal{F}^* c^n)^* y \to F'(\mathcal{F}^* c^\star)^* y \quad \text{for all } y \ , \quad \text{and} \tag{5.6}$$

$$\|F'(\mathcal{F}^* c) - F'(\mathcal{F}^* c')\| \leq L C_2 \|c - c'\| \ . \tag{5.7}$$

These conditions are essentially necessary to establish weak convergence. If F is not equipped with conditions (5.5)-(5.7) as an operator from $X \to Y$, this can be achieved by assuming more regularity of x, i.e. changing the domain of F a little (hoping that the underlying application still fits with modified setting). To this end, we then assume that there exists a function space X^s, and a compact embedding operator $i^s : X^s \to X$. Then we may consider $\tilde{F} = F \circ i^s : X^s \longrightarrow Y$. Lipschitz regularity is preserved. Moreover, if now $x^n \stackrel{w}{\to} x^\star$ in X^s, then $x^n \to x^\star$ in X and, moreover, $\tilde{F}'(x^n) \to \tilde{F}'(x^\star)$ in the operator norm. This argument applies to arbitrary nonlinear continuous and Fréchet differentiable operators $F : X \to Y$ with continuous Lipschitz derivative as long as a function space X^s with compact embedding i^s into X is available.

At a first glance the made assumptions on F might seem to be somewhat restrictive. But compared to usually made assumptions in nonlinear inverse problems they are indeed reasonable and are fulfilled by numerous applications.

All what follows in the remaining section can be comprehensively retraced (including all proofs) in [46].

5.1 Properties of the Surrogate Functional

By the definition of J^s_α in (5.3) it is not clear whether the functional is positive definite or even bounded from below. This will be the case provided the constant C is chosen properly.

For given $\alpha > 0$ and c^0 we define a ball $K_r := \{c \in \ell_2 : \Psi(Bc) \leq r\}$, where the radius r is given by

$$r := \frac{\|y^\delta - F(\mathcal{F}^* c^0)\|^2_Y + 2\alpha\Psi(Bc^0)}{2\alpha}. \tag{5.8}$$

This obviously ensures $c^0 \in K_r$. Furthermore, we define the constant C by

$$C := 2C_2 \max\left\{ \left(\sup_{c \in K_r} \|F'(\mathcal{F}^* c)\| \right)^2, L\sqrt{J_\alpha(c^0)} \right\}, \tag{5.9}$$

where L is the Lipschitz constant of theFrechét derivative of F and C_2 the upper frame bound in (3.6). We assume that c^0 was chosen such that $r < \infty$ and $C < \infty$.

Lemma 5.1. *Let r and C be chosen by* (5.8), (5.9). *Then, for all $c \in K_r$,*

$$C\|c - c^0\|_{\ell_2}^2 - \|F(\mathcal{F}^*c) - F(\mathcal{F}^*c^0)\|_Y^2 \geq 0$$

and thus, $J_\alpha(c) \leq J_\alpha^s(c, c^0)$.

In our iterative approach, this property carries over to all of the iterates.

Proposition 5.2. *Let c^0, α be given and r, C be defined by* (5.8), (5.9). *Then the functionals $J_\alpha^s(c, c^n)$ are bounded from below for all $c \in \ell_2$ and all $n \in \mathbb{N}$ and have thus minimizers. For the minimizer c^{n+1} of $J_\alpha^s(c, c^n)$ holds $c^{n+1} \in K_r$.*

The proof of the latter Proposition 5.2 directly yields

Corollary 5.3. *The sequences $(J_\alpha(c^n))_{n\in\mathbb{N}}$ and $(J_\alpha^s(c^{n+1}, c^n))_{n\in\mathbb{N}}$ are non-increasing.*

5.2 Minimization of the Surrogate Functionals

To derive an algorithm that approximates a minimizer of (5.1), we elaborate the necessary condition.

Lemma 5.4. *The necessary condition for a minimum of $J_\alpha^s(c, a)$ is given by*

$$0 \in -\mathcal{F}F'(\mathcal{F}^*c)^*(y^\delta - F(\mathcal{F}^*a)) + C\,c - C\,a + \alpha B^*\partial\Psi(Bc) \ . \tag{5.10}$$

This result can be achieved when introducing the functional Θ via the relation $v \in \partial\Theta(c) \Leftrightarrow Bv \in \partial\Psi(Bc)$; then one obtaines in the notion of subgradients,

$$\partial J_\alpha^s(c, a) = -2\mathcal{F}F'(\mathcal{F}^*c)^*(y^\delta - F(\mathcal{F}^*a)) + 2C\,c - 2C\,a + 2\alpha\partial\Theta(c) \ .$$

Lemma 5.5. *Let $M(c, a) := \mathcal{F}F'(\mathcal{F}^*c)^*(y^\delta - F(\mathcal{F}^*a))/C + a$. The necessary condition* (5.10) *can then be casted as*

$$c = \frac{\alpha}{C}B^*\left(I - P_{\mathcal{C}}\right)\left(\frac{C}{\alpha}BM(c, a)\right) \ , \tag{5.11}$$

where $P_{\mathcal{C}}$ is an orthogonal projection onto a convex set $\mathcal{C}$.

To verify Lemma 5.5, one has to establish the relation between Ψ and $\mathcal{C}$. To this end, we consider the Fenchel or so–called dual functional of Ψ, which we will denote by Ψ^*. For a functional $\Psi : X \to \overline{\mathcal{R}}$, the dual function $\Psi^* : \mathcal{X}^* \to \overline{\mathcal{R}}$ is defined by

$$\Psi^*(x^*) = \sup_{x\in X}\{\langle x^*, x\rangle - \Psi(x)\} \ .$$

Since we have assumed Ψ to be a positive and one homogeneous functional, there exists a convex set $\mathcal{C}$ such that Ψ^* is equal to the indicator function $\chi_{\mathcal{C}}$ over $\mathcal{C}$. Moreover, in a Hilbert space setting, we have total duality between convex sets and positive and one homogeneous functionals, i.e. $\Psi = (\chi_{\mathcal{C}})^*$.

Consequently, with the shorthand notation $M(c, a)$ we may rewrite (5.10),

$$B\frac{M(c,a) - c}{\frac{\alpha}{C}} \in \partial\Psi(Bc) ,$$

and thus, by convex analysis standard arguments,

$$\frac{C}{\alpha}Bc \in \frac{C}{\alpha}\partial\Psi^* \left(B\frac{M(c,a) - c}{\frac{\alpha}{C}}\right) .$$

In order to have an expression by means of projections, we expand the latter formula as follows

$$\begin{aligned} B\frac{M(c,a)}{\frac{\alpha}{C}} &\in B\frac{M(c,a) - c}{\frac{\alpha}{C}} + \frac{C}{\alpha}\partial\Psi^* \left(B\frac{M(c,a) - c}{\frac{\alpha}{C}}\right) \\ &= \left(I + \frac{C}{\alpha}\partial\Psi^*\right)\left(B\frac{M(c,a) - c}{\frac{\alpha}{C}}\right) , \end{aligned}$$

which is equivalent to

$$\left(I + \frac{C}{\alpha}\partial\Psi^*\right)^{-1}\left(B\frac{M(c,a)}{\frac{\alpha}{C}}\right) = B\frac{M(c,a) - c}{\frac{\alpha}{C}} .$$

Again, by standard arguments, (for more details, see [46])it is known that $\left(I + \frac{C}{\alpha}\partial\Psi^*\right)^{-1}$ is nothing than the orthogonal projection onto a convex set $\mathcal{C}$, and hence the assertion (5.11) follows.

Lemma 5.5 states that for minimizing (5.3) we need to solve the fixed point equation (5.11). To this end, we introduce the associated fixed point map $\Phi_{\alpha,\mathcal{C}}$ with respect to some α and $\mathcal{C}$, i.e.

$$\Phi_{\alpha,\mathcal{C}}(c,a) := \frac{\alpha}{C}B^*(I - P_{\mathcal{C}})\left(B\frac{M(c,a)}{\frac{\alpha}{C}}\right).$$

In order to ensure contractivity of $\Phi_{\alpha,\mathcal{C}}$ for some generic a we need two standard properties of convex sets, see [7].

Lemma 5.6. *Let K be a closed and convex set in some Hilbert space X, then for all $u \in X$ and all $k \in K$ the inequality $\langle u - P_K u, k - P_K u\rangle \leq 0$ holds true.*

Lemma 5.7. *Let K be a closed and convex set, then for all $u, v \in X$ the inequality*

$$\|u - v - (P_K u - P_K v)\| \leq \|u - v\|$$

holds true.

Thanks to Lemmata 5.6 and 5.7 we obtain

Lemma 5.8. *The mapping $I - P_{\mathcal{C}}$ is non–expansive.*

The latter statement provides contractivity of $\Phi_{\alpha,\mathcal{C}}(\cdot, a)$.

Lemma 5.9. *The operator $\Phi_{\alpha,\mathcal{C}}(\cdot, a)$ is a contraction, i.e.*

$$\|\Phi_{\alpha,\mathcal{C}}(c, a) - \Phi_{\alpha,\mathcal{C}}(\tilde{c}, a)\|_{\ell_2} \leq q\|c - \tilde{c}\|_{\ell_2} \quad \textit{if} \quad q := \frac{C_2 L}{C}\sqrt{J_\alpha(a)} < 1 \,.$$

This consequently leads to

Proposition 5.10. *The fixed point map $\Phi_{\alpha,\mathcal{C}}(c, c^n)$ that is applied in* (5.11) *to compute c is due to definition* (5.9) *for all $n = 0, 1, 2, \ldots$ and all $\alpha > 0$ and $\mathcal{C}$ a contraction.*

The last proposition guarantees convergence towards a critical point of $J^s_\alpha(\cdot, c^n)$. This can be sharpened.

Proposition 5.11. *The necessary equation* (5.11) *for a minimum of the functional $J^s_\alpha(\cdot, c^n)$ has a unique fixed point, and the fixed point iteration converges towards the minimizer.*

By assuming more regularity on F, the latter statement can be improved a little.

Proposition 5.12. *Let F be a twice continuously differentiable operator. Then the functional $J^s_\alpha(\cdot, c^n)$ is strictly convex.*

5.3 Convergence Properties

Within this section we establish convergence properties of $(c^n)_{n\in\mathbb{N}}$. In particular, we show that $(c^n)_{n\in\mathbb{N}}$ converges strongly towards a critical point of J_α.

Lemma 5.13. *The sequence of iterates $(c^n)_{n\in\mathbb{N}}$ has a weakly convergent subsequence.*

This is an immediate consequence of Proposition 5.2, in which we have shown that for $n = 0, 1, 2, \ldots$ the iterates c^n remain in K_r. Moreover, with the help of Corollary 5.3, we observe the following lemma that is essentially used in the convergence proof.

Lemma 5.14. *For the iterates c^n holds* $\lim_{n\to\infty} \|c^{n+1} - c^n\|_{\ell_2} = 0$.

To arrive at weak convergence, we need the following preliminary lemmatas. They state properties involving the general constraint Θ. To achieve strong convergence the analysis is limited to the class of constraints given by weighted ℓ_p norms.

Lemma 5.15. *Let Θ be a convex and weakly lower semi–continuous functional. For sequences $v^n \to v$ and $c^n \stackrel{w}{\to} c$, assume $v^n \in \partial\Theta(c^n)$ for all $n \in \mathbb{N}$. Then, $v \in \partial\Theta(c)$.*

Thanks to last Lemma 5.15 we therefore have weak convergence.

Lemma 5.16. *Every subsequence of $(c^n)_{n\in\mathbb{N}}$ has a weakly convergent subsequence $(c^{n_l})_{l\in\mathbb{N}}$ with weak limit $c^\star_\alpha$ that satisfies the necessary condition for a minimizer of J_α,*

$$\mathcal{F}F'(\mathcal{F}^* c^\star_\alpha)^*(y^\delta - F(\mathcal{F}^* c^\star_\alpha)) \in \alpha\partial\Theta(c^\star_\alpha) \; . \tag{5.12}$$

Lemma 5.17. *Let $(c^{n_l})_{l\in\mathbb{N}} \subset (c^n)_{n\in\mathbb{N}}$ with $c^{n_l} \stackrel{w}{\to} c^\star_\alpha$. Then, $\lim_{l\to\infty} \Theta(c^{n_l}) = \Theta(c^\star_\alpha)$*

Combining the previous lemmatas and restricting the constraints to weighted ℓ_p norms, we can achieve strong convergence of the subsequence $(c^{n_l})_{l\in\mathbb{N}}$.

Theorem 5.18. *Let $(c^{n_l})_{l\in\mathbb{N}} \subset (c^n)_{n\in\mathbb{N}}$ with $c^{n_l} \stackrel{w}{\to} c^\star_\alpha$. Assume, moreover, that*

$$\Theta(c) = \Psi(c) = \left(\sum_j w_j |c_j|^p \right)^{1/p}$$

with $w_j \geq r > 0$ and $1 \leq p \leq 2$. Then the subsequence $(c^{n_l})_{l\in\mathbb{N}}$ converges also in norm.

In principle, the limits of different convergent subsequences of c^n may differ. Let $c^{n_l} \to c^\star_\alpha$ be a subsequence of c^n, and let $c^{n_l\prime}$ the predecessor of c^{n_l} in c^n, i.e. $c^{n_l} = c^i$ and $c^{n_l\prime} = c^{i-1}$. Then we observe, $J^s_\alpha(c^{n_l}, c^{n_l\prime}) \to J_\alpha(c^\star_\alpha)$. Moreover, as we have $J^s_\alpha(c^{n+1}, c^n) \leq J^s_\alpha(c^n, c^{n-1})$ for all n, it turns out that the value of the Tikhonov functional for every limit $c^\star_\alpha$ of a convergent subsequence remains the same, i.e. $J_\alpha(c^\star_\alpha) = const$.

We summarize our findings and give a simple criterion that ensures strong convergence of the whole sequence $(c^n)_{n\in\mathbb{N}}$ towards a critical point of J_α.

Theorem 5.19. *Assume that there exists at least one isolated limit $c^\star_\alpha$ of a subsequence c^{n_l} of c^n. Then $c^n \to c^\star_\alpha$ as $n \to \infty$. The accumulation point $c^\star_\alpha$ is a minimizer for the functional $J^s_\alpha(\cdot, c^\star_\alpha)$ and fulfills the necessary condition for a minimizer of J_α.*

Moreover, we obtain, $J^s_\alpha(c^\star_\alpha + h, c^\star_\alpha) \geq J^s_\alpha(c^\star_\alpha, c^\star_\alpha) + \frac{C}{2}\|h\|^2$ and with Lemma 5.12 the second assertion in the theorem can be shown. The first assertion of the theorem can be directly taken from [44].

As a summary of the above reasoning we suggest the following implementation of the proposed Tikhonov-based projection iteration.

Iterated (generalized) Shrinkage

for nonlinear ill-posed and inverse problems with sparsity constraints

Given	operator F, its derivative F', matrix B, data y^δ, some initial guess c^0, and $\alpha > 0$
Initialization	$K_r = \{c \in \ell_2 : \Psi(Bc) \le r\}$ with $r = J_\alpha(c^0)/(2\alpha)$, $C = 2C_2 \max\{\sup_{c \in K_r} \|F'(\mathcal{F}^* c)\|^2, L\sqrt{J_\alpha(c^0)}\}$
Iteration	**for** $n = 0, 1, 2, \ldots$ **until** a preassigned precision / maximum number of iterations **1.** $c^{n+1} = \frac{\alpha}{C} B^*(I - P_{\mathcal{C}}) \left(\frac{C}{\alpha} B\, M(c^{n+1}, c^n)\right)$ by fixed point iteration, and where $M(c^{n+1}, c^n) = c^n + \frac{1}{C} \mathcal{F} F'(\mathcal{F}^* c^{n+1})^* (y^\delta - F(\mathcal{F}^* c^n))$ **end**

5.4 Application of Sparse Recovery to SPECT

This section is devoted to the application of the developed theory to a sparse recovery problem in the field of single photon emission computed tomography (SPECT). SPECT is a medical imaging technique where one aims to reconstruct a radioactivity distribution f from radiation measurements outside the body. The measurements are described by the attenuated Radon transform (ATRT)

$$y = A(f, \mu)(s, \omega) = \int_{\mathbb{R}} f(s\omega^\perp + t\omega) e^{-\int_t^\infty \mu(s\omega^\perp + r\omega) dr} dt \,. \tag{5.13}$$

As the measurements depend on the (usually also unknown) density distribution μ of the tissue, we have to solve a nonlinear problem in (f, μ). A throughout analysis of the nonlinear ATRT was presented by Dicken [22], and several approaches for its solution were proposed in [4, 29, 56, 57, 43, 40, 41, 42]. If the ATRT operator is considered with

$$D(A) = H_0^{s_1}(\Omega) \times H_0^{s_2}(\Omega) \,,$$

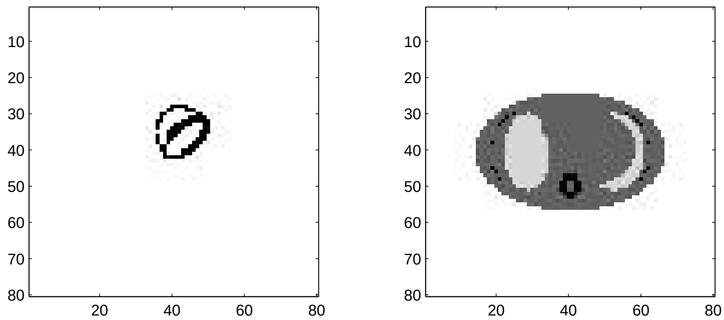

Figure 1. Activity function f^* (left) and attenuation function μ^* (right). The activity function models a cut through the heart.

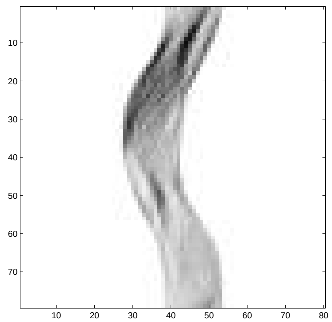

Figure 2. Generated data $g(s, \omega) = A(f^*, \mu^*)(s, \omega)$.

where $H_0^s(\Omega)$ denotes a Sobolev space over a bounded area Ω with zero boundary conditions and smoothness s, then the operator is twice continuous Fréchet differentiable with Lipschitz continuous first derivative, if $s_1,\ s_2$ are chosen large enough. A possible choice for these parameters that also reflects the smoothness properties of activity and density distribution is $s_1 > 4/9$ and $s_2 = 1/3$. For more details we refer to [41, 21]. Additionally, it has been shown that conditions (5.5), (5.7) hold [39]. For our test computations, we will use the so called MCAT – phantom [51], see Figure 1. Both functions were given as 80×80 pixel images. The sinogram data was gathered on 79 angles, equally spaced over 360 degree, and 80 samples. The sinogram belonging to the MCAT phantom is shown in Figure 2.

At first, we have to choose the underlaying frame or basis on which we put the sparsity constraint. Since a wavelet expansion might sparsely represent images/functions (better than pixel basis), we have chosen a wavelet basis (here Daubechies wavelets of

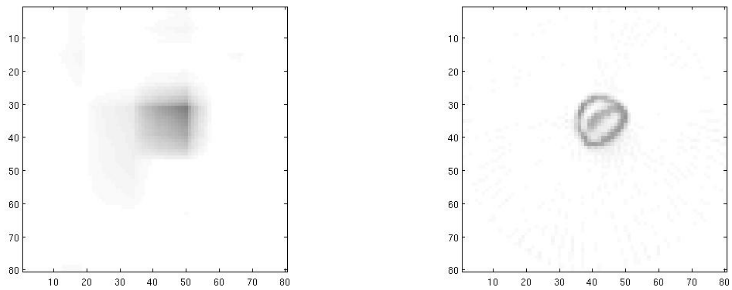

Figure 3. Reconstructions with 5% noise and $\alpha = 350$: sparsity constraint (left) and Hilbert space constraint (right).

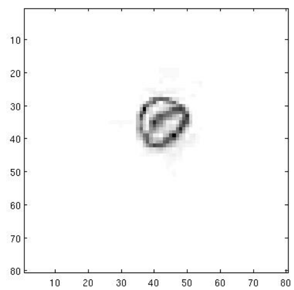

Figure 4. Reconstruction with sparsity constraint and 5% noise. The regularization parameter ($\alpha = 5$) was chosen such that $\|y^\delta - A(f^*, \mu^*)\| \approx 2\delta$

order two) to represent (f, μ), i.e.

$$(f,\mu) = \left(\sum_k c(f)_k \phi_{0,k} + \sum_{j \geq 0, i, k} d(f)^i_{j,k} \psi^i_{j,k} \; , \; \sum_k c(\mu)_k \phi_{0,k} + \sum_{j \geq 0, i, k} d(\mu)^i_{j,k} \psi^i_{j,k} \right) .$$

For more details we refer the reader to [12]. For our implementation we have chosen $B = I$ and $\Psi(\cdot) = \|\cdot\|_{\ell_1}$. As an observation, the speed of convergence depends heavily on the choice of the constant C in (5.3). According to our convergence analysis, it has to be chosen reasonably large. However, a large C speeds up the convergence of the inner iteration, but decreases the speed of convergence of the outer iteration. In our example, we needed only 2-4 inner iteration, but the outer iteration required about 5000 iterations. As the minimization in the quadratic case needed much less iterations, this suggests that the speed of convergence also increases with p.

According to (5.2), the functional Ψ will always have a bigger value than $\| \cdot \|_{\ell_2}$. If $\Psi(c)$ is not too large, then it will also dominate $\|c\|^2_{\ell_2}$, which also represents the classical L_2-norm, and we might conclude that reconstructions with the classical quadratic

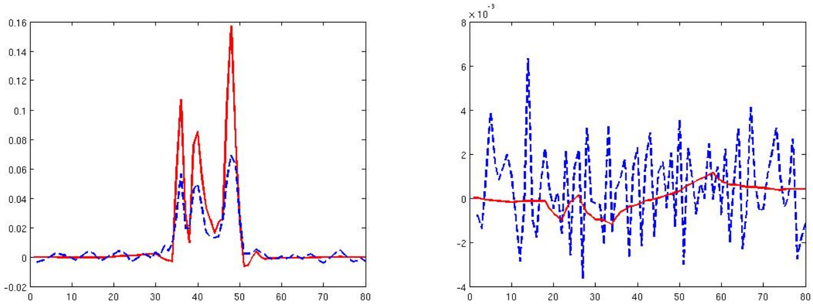

Figure 5. Values of the reconstructed activity function through the heart (left) and well below the heart (right). Solid line: reconstruction with sparsity constraint, dashed line: quadratic Hilbert space penalty

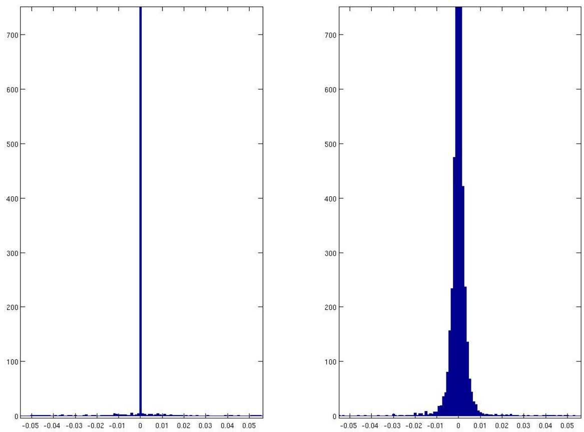

Figure 6. Histogram plot of the wavelet coefficient of the reconstructions. Left: sparsity constraint, Right: quadratic Hilbert space constraint.

Hilbert space constraint and sparsity constraint will not give comparable results if the same regularization parameter is used. As Ψ is dominant, we expect a smaller (optimal) regularization parameter in the case of the penalty term Ψ. This is confirmed by our first test computations: Figure 3 shows the reconstructions from noisy data where the regularization parameter was chosen as $\alpha = 350$. The reconstruction with the quadratic Hilbert space penalty (we have used the L_2 norm) is already quite good, whereas the reconstruction for the sparsity constraint is still far off. In fact, if we consider Morozov's discrepancy principle, then the regularization parameter in the quadratic case has been chosen optimal, as we observe

$$\|y^\delta - A(f_\alpha^\delta, \mu_\alpha^\delta)\| \approx 2\delta \; .$$

To obtain a reasonable basis for comparison, we adjusted the regularization parameter α such that the residual had also a magnitude of 2δ in the sparsity case, which occurred for $\alpha = 5$. The reconstruction can be seen in Figure 4.

A visual inspection shows that the reconstruction with sparsity constraint yields much sharper contours. In particular, the absolute values of f in the heart are higher in the sparsity case, and the artifacts are not as bad as in the quadratic constraint case, as can be seen in Figure 5. It shows a plot of the values of the activity function for both reconstructions along a row in the image in Figures 3 and 4 respectively. The left graph shows the values at a line that goes through the heart, and right graph shows the values along a line well outside the heart, where only artifacts occur. Clearly, both reconstructions are different, but it certainly needs much more computations in order to decide in which situations a sparsity constraint has to be preferred. A histogram plot of the wavelet coefficients for both reconstructions shows that the reconstruction with sparsity constraint has much more small coefficients - it is, as we did expect, a sparse reconstruction, see Figure 6.

6 Projected Accelerated Steepest Descent for Nonlinear Ill-Posed Problems

In the previous section we have discussed the iterated generalized shrinkage method given by

$$c^{n+1} = \frac{\alpha}{C} B^*(I - P_{\mathcal{C}}) \left(\frac{C}{\alpha} B \left\{ c^n + \mathcal{F}F'(\mathcal{F}^* c^{n+1})^*(y^\delta - F(\mathcal{F}^* c^n))/C \right\} \right) \; .$$

For special choices of Ψ (e.g. $\Psi(c) = \|c\|_{\ell_1}$), this iteration then allows due its simple nature an easy to implement recovery algorithm. But the convergence is rather slow and does not change substantially through different choices of Ψ. One first serious step to accelerate such types of iterations (but for linear problems) was suggested in [15], in which the authors “borrowed a leaf” from standard linear steepest descent methods by using an adaptive step length. In addition to this, the authors concluded from a detailed analysis of the characteristic dynamics of the iterated soft-shrinkage that the algorithm converges initially relatively fast, then it overshoots the ℓ_1 penalty, and it takes very long to re-correct back. The proposed way to circumvent this “external” detour is to force the iterates to remain within a particular ℓ_1 ball $B_R := \{c \in \ell_2; \|c\|_{\ell_1} \leq R\}$. This has led to the constrained optimization problem

$$\min_{c \in B_R} \|A\mathcal{F}^* c - y^\delta\|^2 \tag{6.1}$$

resulting in a significantly different proceeding. The shrinkage operation is replaced by a projection P_{B_R} (where the projection $\mathrm{P}_{\mathcal{C}}(c)$ is defined for any closed convex set $\mathcal{C}$ and any c as the unique point in $\mathcal{C}$ for which the ℓ_2 distance to c is minimal) leading for properly chosen $\gamma > 0$ to the following iteration,

$$c^{n+1} = \mathrm{P}_{B_R}(c^n + \gamma \mathcal{F} A^*(y^\delta - A\mathcal{F}^* c^n)). \tag{6.2}$$

However, the speed of convergence remained very slow. Therefore, as mentioned above, the authors suggested to introduce an adaptive "descent parameter" $\gamma^n > 0$ in each iteration yielding

$$c^{n+1} = \mathrm{P}_{B_R}(c^n + \gamma^n \mathcal{F} A^*(y^\delta - A\mathcal{F}^* c^n)) . \tag{6.3}$$

The authors of [15] referred to this modified algorithm as the *projected gradient iteration* or the *projected steepest descent method*. They have determined how large one can choose the successive γ^n and have shown weak as well as strong convergence of the method (with and without acceleration). Alternative approaches for sparse recovery that are closely related to the introduced method are the schemes presented in [32] and [55]. The analysis in [55] is limited to finite dimensions and the strategy provided in [32] is suited for linear inverse problems. The principle there is to reformulate the minimization problem as a bounded constrained quadratic program, and then apply iterative project gradient iterations.

In this section we show that iteration (6.3) (and also more general formulations) can be directly extended to the nonlinear situation resulting in

$$c^{n+1} = \mathrm{P}_{B_R}(c^n + \gamma^n \mathcal{F} F'(\mathcal{F}^* c^{n+1})^*(y - F(\mathcal{F}^* c^n))) . \tag{6.4}$$

Again, as in the previous section, weak as well as strong convergence can only be achieved, if F is equipped with conditions (5.5)-(5.7). We also assume twice continuous Fréchet differentiability of F. But note that at the cost of more technicalities most of the results can also be achieved if F is only one time Fréchet differentiable.

Another issue that is of great importance but was neither considered in [15] nor somewhere else is to verify regularizing properties of (6.4). Elaborations on this topic, however, are not provided so far. Nevertheless, we wish to briefly mention the theory that is still provided in the literature, which is so far unfortunately limited to linear problems, see, e.g., [23, Section 5.4]. Therefore, the concepts summarized in [23] not directly apply here and need to be extended. In any case, the question arises whether the convex constraint stabilize the problem or if it is still necessary to regularize the inverse problem. In general it seems to be meaningful to assume ill-posedness. Therefore, we need to introduce an additional stabilization. The iteration (6.4) can be viewed as iteration scheme to approach the B_R*-best-approximate solution* $c_R^\dagger$, which we define as the minimizer of $\|F(\mathcal{F}^*c) - y\|^2$ on B_R, i.e.

$$\begin{aligned} \|F(\mathcal{F}^*c_R^\dagger) - y\| &= \inf_c\{\|F(\mathcal{F}^*c) - y\|,\ c \in B_R\} \text{ and} \\ \|c_R^\dagger\| &= \min\{\|c\|,\ \|F(\mathcal{F}^*c) - y\| = \|F(\mathcal{F}^*c_R^\dagger) - y\| \text{ and } c \in B_R\}\ . \end{aligned}$$

Since $c_R^\dagger \in B_R$, it is natural to require that the regularized solutions are in B_R as well. If $c^\dagger$ denotes the generalized solution of the unconstrained problem and if $c_R^\dagger = c^\dagger$, then all "standard results" concerning stability, convergence, and convergence rates hold also for the constrained case. If $c_R^\dagger \neq c^\dagger$, one might select a different regularization method, e.g.,

$$\min_{c \in B_R} \|F(\mathcal{F}^*c) - y\|^2 + \eta\|c\|^2\ ,$$

for some $\eta > 0$.

6.1 Preleminaries

Once a frame is selected for X, the computation of a solution x translates into finding a corresponding sequence $c \in \ell_2(\Lambda)$. Hence, the operator under consideration can be written as $F \circ \mathcal{F}^* : \ell_2(\Lambda) \to Y$. Thus, for the ease of notation we write in the remaining section (if not misleadingly used) only F instead of $F \circ \mathcal{F}^*$.

Before analyzing the proposed projected steepest descent (6.4), we provide some analysis of ℓ_2 projections onto ℓ_1 balls. The listed properties can be retraced in [15, 53], from where they are partially taken, or to some extent in [18, 19].

Lemma 6.1. $\forall a \in \ell_2(\Lambda), \forall \tau > 0 : \|\mathrm{S}_\tau(a)\|_1$ *is a piecewise linear, continuous, decreasing function of* τ*; moreover, if* $a \in \ell_1(\Lambda)$ *then* $\|\mathrm{S}_0(a)\|_1 = \|a\|_1$ *and* $\|\mathrm{S}_\tau(a)\|_1 = 0$ *for* $\tau \geq \max_i |a_i|$.

Lemma 6.2. *If* $\|a\|_1 > R$*, then the* ℓ_2 *projection of* a *on the* ℓ_1 *ball with radius* R *is given by* $\mathrm{P}_{B_R}(a) = \mathrm{S}_\mu(a)$*, where* μ *(depending on* a *and* R*) is chosen such that* $\|\mathrm{S}_\mu(a)\|_1 = R$*. If* $\|a\|_1 \leq R$ *then* $\mathrm{P}_{B_R}(a) = \mathrm{S}_0(a) = a$.

Lemma 6.1 and 6.2 provide a simple recipe for computing the projection $\mathrm{P}_{B_R}(a)$. First, sort the absolute values of the components of a (an $\mathcal{O}(m \log m)$ operation if $\#\Lambda = m$ is finite), resulting in the rearranged sequence $(a_l^*)_{l=1,\ldots,m}$, with $a_l^* \geq a_{l+1}^* \geq 0; \forall l$. Next, perform a search to find k such that

$$\|\mathrm{S}_{a_k^*}(a)\|_1 = \sum_{l=1}^{k-1}(a_l^* - a_k^*) \leq R < \sum_{l=1}^{k}(a_l^* - a_{k+1}^*) = \|\mathrm{S}_{a_{k+1}^*}(a)\|_1\ .$$

The complexity of this step is again $\mathcal{O}(m \log m)$. Finally, set $\nu := k^{-1}(R - \|S_{a_k^*}(a)\|_1)$, and $\mu := a_k^* - \nu$. Then

$$\begin{aligned} \|\mathrm{S}_\mu(a)\|_1 &= \sum_{i\in\Lambda} \max(|a_i| - \mu, 0) = \sum_{l=1}^{k} (a_l^* - \mu) \\ &= \sum_{l=1}^{k-1} (a_l^* - a_k^*) + k\nu = \|\mathrm{S}_{a_k^*}(a)\|_1 + k\nu = R \,. \end{aligned}$$

In addition to the above statements, also the still provided Lemmata 5.6 (setting $K = B_R$), 5.7, and 5.8 apply to P_{B_R} and allow therewith the use of several standard arguments of convex analysis.

6.2 Projected Steepest Descent and Convergence

We have now collected some facts on the projector P_{B_R} and on convex analysis issues that allow for convergence analysis of the projected steepest descent method defined in (6.3). In what follows, we essentially proceed as in [15]. But as we shall see, several serious technical changes (including also a weakening of a few statements) but also significant extensions of the nice analysis provided in [15] need to be made. For instance, due to the nonlinearity of F, several uniqueness statements proved in [15] do not carry over in its full glory. Nevertheless, the main propositions on *weak* and *strong convergence* can be achieved (of course, at the cost of involving much more technicalities).

First, we derive the necessary condition for a minimizer of $\boldsymbol{D}(c) := \|F(c) - y\|^2$ on B_R.

Lemma 6.3. *If the vector $\tilde{c}_R \in \ell_2$ is a minimizer of $\boldsymbol{D}(c)$ on B_R then for any $\gamma > 0$,*

$$\mathrm{P}_{B_R}(\tilde{c}_R + \gamma F'(\tilde{c}_R)^*(y - F(\tilde{c}_R))) = \tilde{c}_R \,,$$

which is equivalent to

$$\langle F'(\tilde{c}_R)^*(y - F(\tilde{c}_R)), w - \tilde{c}_R \rangle \leq 0, \quad for\ all\ w \in B_R \,.$$

This result essentially relies on the Fréchet differentiability of F (see, e.g., [59, 58]) and summarizes the following reasoning.

With the help of the first order Taylor expansion given by

$$F(c+h) = F(c) + F'(c)h + R(c,h) \quad \text{with} \quad \|R(c,h)\| \leq \frac{L}{2}\|h\|^2$$

one has for the minimizer $\tilde c_R$ of $\boldsymbol{D}$ on B_R and all $w \in B_R$ and all $t \in [0,1]$

$$\begin{aligned}
\boldsymbol{D}(\tilde c_R) \quad &\le \quad = \boldsymbol{D}(\tilde c_R + t(w-\tilde c_R)) = \|F(\tilde c_R + t(w-\tilde c_R)) - y\|^2 \\
&= \quad \|F(\tilde c_R) - y + F'(\tilde c_R)t(w-\tilde c_R) + R(\tilde c_R, t(w-\tilde c_R))\|^2 \\
&= \quad \boldsymbol{D}(\tilde c_R) + 2\langle F'(\tilde c_R)^*(F(\tilde c_R)-y), t(w-\tilde c_R)\rangle \\
&\qquad +2\langle F(\tilde c_R)-y, R(\tilde c_R, t(w-\tilde c_R))\rangle \\
&\qquad\quad +\|F'(\tilde c_R)t(w-\tilde c_R) + R(\tilde c_R, t(w-\tilde c_R))\|^2 \ .
\end{aligned}$$

This implies

$$\langle F'(\tilde c_R)^*(y - F(\tilde c_R)), w - \tilde c_R\rangle \le 0 \ ,$$

and therefore, for all $\gamma > 0$,

$$\langle \tilde c_R + \gamma F'(\tilde c_R)^*(y - F(\tilde c_R)) - \tilde c_R, w - \tilde c_R\rangle \le 0 \ ,$$

which verifies the assertion.

Lemma 6.3 provides just a necessary condition for a minimizer $\tilde c_R$ of $\boldsymbol{D}$ on B_R. The minimizers of $\boldsymbol{D}$ on B_R need not be unique. Nevertheless, we have

Lemma 6.4. *If $\tilde c, \tilde{\tilde c} \in B_R$, if $\tilde c$ minimizes $\boldsymbol{D}$ and if $\tilde c - \tilde{\tilde c} \in \ker F'(w)$ for all $w \in B_R$ then $\tilde{\tilde c}$ minimizes $\boldsymbol{D}$ as well.*

In what follows we elaborate the convergence properties of (6.4). In a first step we establish weak convergence and in a second step we extend weak to strong convergence. To this end, we have to specify the choice of γ^n. At first, we introduce a constant r,

$$r := \max\{2 \sup_{c \in B_R} \|F'(c)\|^2 \ , \ 2L\sqrt{\boldsymbol{D}(c^0)}\} \ , \tag{6.5}$$

where c^0 denotes a initial guess for the solution to be reconstructed. One role of the constant r can be seen in the following estimate which is possible by the first order Taylor expansion of F,

$$\|F(c^{n+1}) - F(c^n)\|^2 \le \sup_{c \in B_R} \|F'(c)\|^2 \|c^{n+1} - c^n\|^2 \le \frac{r}{2}\|c^{n+1} - c^n\|^2 \ .$$

We define now with the help of (6.5) a sequence of real numbers which we denote by β^n that specifies the choice γ^n by setting $\gamma^n = \beta^n / r$ (as we shall see later in this section).

Definition 6.5. We say that the sequence $(\beta^n)_{n\in\mathbb{N}}$ satisfies Condition (B) with respect to the sequence $(c^n)_{n\in\mathbb{N}}$ if there exists n_0 such that:

(B1) $\quad \bar\beta := \sup\{\beta^n; n \in \mathbb{N}\} < \infty \quad \text{and} \quad \inf\{\beta^n; n \in \mathbb{N}\} \ge 1$

(B2) $\quad \beta^n \|F(c^{n+1}) - F(c^n)\|^2 \le \frac{r}{2}\|c^{n+1} - c^n\|^2 \quad \forall n \ge n_0$

(B3) $\quad \beta^n L\sqrt{\boldsymbol{D}(c^n)} \le \frac{r}{2}$.

By condition (B1) we ensure

$$\|F(c^{n+1}) - F(c^n)\|^2 \le \beta^n \|F(c^{n+1}) - F(c^n)\|^2 \,.$$

The idea of adding condition (B2) is to find the largest number $\beta^n \ge 1$ such that

$$0 \le -\|F(c^{n+1}) - F(c^n)\|^2 + \frac{r}{2\beta^n}\|c^{n+1} - c^n\|^2$$

is as small as possible. The reason can be verified below in the definition of the surrogate functional Φ_β in Lemma 6.6. The goal is to ensure that Φ_{β^n} is not too far off $\boldsymbol{D}(c^n)$. The additional restriction (B3) is introduced to ensure convexity of Φ_{β^n} and convergence of the fixed point map Ψ in Lemma 6.7 (as we will prove below).

Because the definition of c^{n+1} involves β^n and vice versa, the inequality (B2) has an implicit quality. In practice, it is not straightforward to pick β^n adequately. This issue will be discussed later in Subsection 6.3.

In the remaining part of this subsection we prove weak convergence of any subsequence of $(c^n)_{n\in\mathbb{N}}$ towards weak limits that fulfill the necessary condition for minimizers of $\boldsymbol{D}$ on B_R.

Lemma 6.6. *Assume F to be twice Fréchet differentiable and $\beta \ge 1$. For arbitrary fixed $c \in B_R$ assume $\beta L\sqrt{\boldsymbol{D}(c)} \le r/2$ and define the functional $\Phi_\beta(\cdot, c)$ by*

$$\Phi_\beta(w,c) := \|F(w) - y\|^2 - \|F(w) - F(c)\|^2 + \frac{r}{\beta}\|w - c\|^2 \,. \tag{6.6}$$

Then there exists a unique $w \in B_R$ that minimizes the restriction to B_R of $\Phi_\beta(w,c)$. We denote this minimizer by $\hat{c}$ which is given by

$$\hat{c} = \mathrm{P}_{B_R}\left(c + \frac{\beta}{r}F'(\hat{c})^*(y - F(c))\right) \,.$$

The essential strategy of the proof goes as follows. First, since F is twice Fréchet differentiable one verifies that $\Phi_\beta(\cdot, c)$ is strictly convex in w. Therefore there exists a unique minimizer $\hat{c}$ and thus we have for all $w \in B_R$ and all $t \in [0,1]$

$$\Phi_\beta(\hat{c}, c) \le \Phi_\beta(\hat{c} + t(w - \hat{c}), c) \,.$$

With the short hand notation $J(\cdot) := \Phi_\beta(\cdot, c)$ it therefore follows that

$$\begin{aligned}
0 \;&\le\; J(\hat{c} + t(w-\hat{c})) - J(\hat{c}) = tJ'(\hat{c})(w - \hat{c}) + \rho(\hat{c}, t(w-\hat{c})) \\
&=\; 2t\langle F(c) - y, F'(\hat{c})(w-\hat{c})\rangle + 2t\frac{r}{\beta}\langle \hat{c} - c, w - \hat{c}\rangle \\
&\qquad + 2\langle F(c) - y, R(\hat{c}, t(w-\hat{c}))\rangle + \frac{r}{\beta}\|t(w-\hat{c})\|^2 \\
&\le\; 2t\left\{\langle F(c) - y, F'(\hat{c})(w-\hat{c})\rangle + \frac{r}{\beta}\langle \hat{c} - c, w - \hat{c}\rangle\right\} \\
&\qquad + t^2\left\{2\frac{r}{2\beta L}\frac{L}{2}\|w - \hat{c}\|^2 + \frac{r}{\beta}\|w-\hat{c}\|^2\right\} \,.
\end{aligned}$$

This implies for all $t \in [0,1]$

$$0 \leq \left\{ \frac{\beta}{r} \langle F(c) - y, F'(\hat{c})(w - \hat{c}) \rangle + \langle \hat{c} - c, w - \hat{c} \rangle \right\} + \frac{3t}{4} \|w - \hat{c}\|^2 .$$

Consequently, we deduce

$$\langle c + \frac{\beta}{r} F'(\hat{c})^*(y - F(c)) - \hat{c}, w - \hat{c} \rangle \leq 0$$

which is equivalent to

$$\hat{c} = \mathrm{P}_{B_R} \left(c + \frac{\beta}{r} F'(\hat{c})^*(y - F(c)) \right)$$

and the assertion is shown.

The unique minimizer $\hat{c}$ is only implicitly given. We propose to apply a simple fixed point iteration to derive $\hat{c}$. The next lemma verifies that the corresponding fixed point map is indeed contractive and can therefore be used.

Lemma 6.7. *Assume $\beta L \sqrt{\boldsymbol{D}(x)} \leq r/2$. Then the map*

$$\Psi(\hat{c}) := \mathrm{P}_{B_R}(c + \beta / r F'(\hat{c})^*(y - F(c)))$$

is contractive and therefore the fixed point iteration $\hat{c}^{l+1} = \Psi(\hat{c}^l)$ converges to a unique fixed point.

The latter Lemma is a consequence of the Lipschitz continuity of F' and the non-expansiveness of P_{B_R}. The last property that is needed to establish convergence is an immediate consequence of Lemma 6.6.

Lemma 6.8. *Assume c^{n+1} is given by*

$$c^{n+1} = \mathrm{P}_{B_R} \left(c^n + \frac{\beta^n}{r} F'(c^{n+1})^*(y - F(c^n)) \right) ,$$

where r is as in (6.5) *and the β^n satisfy Condition* (B) *with respect to $(c^n)_{n\in\mathbb{N}}$, then the sequence $\boldsymbol{D}(c^n)$ is monotonically decreasing and* $\lim_{n\to\infty} \|c^{n+1} - c^n\| = 0$.

Now we have all ingredients for the convergence analysis together. Since for all the iterates we have by definition $c^n \in B_R$, we automatically have $\|c^n\|_2 \leq R$ for all $n \in \mathbb{N}$. Therefore, the sequence $(c^n)_{n\in\mathbb{N}}$ must have weak accumulation points.

Proposition 6.9. *If $c^\star$ is a weak accumulation point of $(c^n)_{n\in\mathbb{N}}$, then it fulfills the necessary condition for a minimum of $\boldsymbol{D}(c)$ on B_R, i.e. for all $w \in B_R$,*

$$\langle F'(c^\star)^*(y - F(c^\star)), w - c^\star \rangle \leq 0 . \tag{6.7}$$

Since this proposition is essential and combines all the above made statements, we give the reasoning and arguments to verify (6.7) in greater detail. Since $c^{n_j} \xrightarrow{w} c^\star$, we have for fixed c and a

$$\langle F'(c)c^{n_j}, a\rangle = \langle c^{n_j}, F'(c)^* a\rangle \longrightarrow \langle c^\star, F'(c)^* a\rangle = \langle F'(c)c^\star, a\rangle$$

and therefore

$$F'(c)c^{n_j} \xrightarrow{w} F'(c)c^\star. \tag{6.8}$$

Due to Lemma 6.8, we also have $c^{n_j+1} \xrightarrow{w} c^\star$. Now we are prepared to show the necessary condition for the weak accumulation point $c^\star$. As the iteration is given by

$$c^{n+1} = \mathrm{P}_{B_R}(c^n + \beta^n/rF'(c^{n+1})^*(y - F(c^n)))\ ,$$

we have

$$\langle c^n + \beta^n/rF'(c^{n+1})^*(y - F(c^n)) - c^{n+1}, w - c^{n+1}\rangle \leq 0 \quad \text{for all } w \in B_R\ .$$

Specializing this inequality to the subsequence $(c^{n_j})_{j\in\mathbb{N}}$ yields

$$\langle c^{n_j} + \beta^{n_j}/rF'(c^{n_j+1})^*(y - F(c^{n_j})) - c^{n_j+1}, w - c^{n_j+1}\rangle \leq 0 \quad \text{for all } w \in B_R\ .$$

Therefore we obtain (due to Lemma 6.8)

$$\limsup_{j\to\infty} \beta^{n_j}/r\langle F'(c^{n_j+1})^*(y - F(c^{n_j})), w - c^{n_j+1}\rangle \leq 0 \quad \text{for all } w \in B_R\ .$$

To the latter inequality we may add

$$\beta^{n_j}/r\langle(-F'(c^{n_j+1})^* + F'(c^{n_j})^*)(y - F(c^{n_j})), w - c^{n_j+1}\rangle$$

and

$$\beta^{n_j}/r\langle F'(c^{n_j})^*(y - F(c^{n_j})), -c^{n_j} + c^{n_j+1}\rangle$$

resulting in

$$\limsup_{j\to\infty} \beta^{n_j}/r\langle F'(c^{n_j})^*(y - F(c^{n_j})), w - c^{n_j}\rangle \leq 0 \quad \text{for all } w \in B_R\ , \tag{6.9}$$

which is possible due to

$$\begin{aligned} &|\langle(-F'(c^{n_j+1})^* + F'(c^{n_j})^*)(y - F(c^{n_j})), w - c^{n_j+1}\rangle| \\ &\quad \leq \quad L\|c^{n_j+1} - c^{n_j}\|\|y - F(c^{n_j})\|\|w - c^{n_j+1}\| \overset{j\to\infty}{\longrightarrow} 0 \end{aligned}$$

and

$$\begin{aligned} &|\langle F'(c^{n_j})^*(y - F(c^{n_j})), -c^{n_j} + c^{n_j+1}\rangle| \\ &\quad \leq \quad \sup_{x\in B_R} \|F'(c)^*\|\|y - F(c^{n_j})\|\|c^{n_j} - c^{n_j+1}\| \overset{j\to\infty}{\longrightarrow} 0\ . \end{aligned}$$

Let us now consider the inner product in (6.9) which we write as

$$\langle F'(c^{n_j})^* y, w - c^{n_j}\rangle - \langle F'(c^{n_j})^* F(c^{n_j}), w - c^{n_j}\rangle \; .$$

For the left summand we have by the weak convergence of $(c^{n_j})_{j\in\mathbb{N}}$ or likewise $(F'(c^\star)c^{n_j})_{j\in\mathbb{N}}$ and the assumption of F, $F'(c^{n_j})^* y \stackrel{j\to\infty}{\longrightarrow} F'(c^\star)^* y$,

$$\begin{aligned}
\langle F'(c^{n_j})^* y, w - c^{n_j}\rangle &= \langle (F'(c^{n_j})^* - F'(c^\star)^* + F'(c^\star)^*) y, w - c^{n_j}\rangle \\
&= \langle F'(c^{n_j})^* y - F'(c^\star)^* y, w - c^{n_j}\rangle + \langle F'(c^\star)^* y, w - c^{n_j}\rangle \\
&\stackrel{j\to\infty}{\longrightarrow} \langle F'(c^\star)^* y, w - c^\star\rangle \\
&= \langle F'(c^\star)^*(y - F(c^\star)), w - c^\star\rangle + \langle F'(c^\star)^* F(c^\star), w - c^\star\rangle \; .
\end{aligned}$$

Therefore (and since $1 \le \beta^{n_j} \le \bar\beta$ and again by the weak convergence of $(c^{n_j})_{j\in\mathbb{N}}$), inequality (6.9) transforms to

$$\begin{aligned}
&\limsup_{j\to\infty} \big[\langle F'(c^\star)^*(y - F(c^\star)), w - c^\star\rangle \\
&\qquad + \langle F'(c^\star)^* F(c^\star), w - c^\star + c^{n_j} - c^{n_j}\rangle - \langle F'(c^{n_j})^* F(c^{n_j}), w - c^{n_j}\rangle\big] \le 0 \\
\Longleftrightarrow& \\
&\limsup_{j\to\infty} \big[\langle F'(c^\star)^*(y - F(c^\star)), w - c^\star\rangle \\
&\qquad + \langle F'(c^\star)^* F(c^\star) - F'(c^{n_j})^* F(c^{n_j}), w - c^{n_j}\rangle\big] \le 0 \\
\Longleftrightarrow& \\
&\langle F'(c^\star)^*(y - F(c^\star)), w - c^\star\rangle \\
&\qquad + \limsup_{j\to\infty} \langle F'(c^\star)^* F(c^\star) - F'(c^{n_j})^* F(c^{n_j}), w - c^{n_j}\rangle \le 0 \; .
\end{aligned}$$

It remains to show that the right summand in (6.10) is for all $w \in B_R$ zero. We have by the assumptions made on F,

$$
\begin{aligned}
&|\langle F'(c^\star)^*F(c^\star) - F'(c^{n_j})^*F(c^{n_j}), w - c^{n_j}\rangle| = \\
&|\langle F'(c^\star)^*F(c^\star) - F'(c^\star)^*F(c^{n_j}) + F'(c^\star)^*F(c^{n_j}) - F'(c^{n_j})^*F(c^{n_j}), w - c^{n_j}\rangle| \\
&\leq \quad |\langle F'(c^\star)^*F(c^\star) - F'(c^\star)^*F(c^{n_j}), w - c^{n_j}\rangle| \\
&\qquad\qquad + |\langle F'(c^\star)^*F(c^{n_j}) - F'(c^{n_j})^*F(c^{n_j}), w - c^{n_j}\rangle| \\
&\leq \quad \sup_{x\in B_R} \|F'(c)\| \|F(c^\star) - F(c^{n_j})\| \|w - c^{n_j}\| \\
&\qquad\qquad + |\langle (F'(c^\star)^* - F'(c^{n_j})^*)(F(c^\star) - F(c^\star) + F(c^{n_j})), w - c^{n_j}\rangle| \\
&\leq \quad \sup_{x\in B_R} \|F'(c)\| \|F(c^\star) - F(c^{n_j})\| \|w - c^{n_j}\| \\
&\qquad\qquad + \|(F'(c^\star)^* - F'(c^{n_j})^*)F(c^\star)\| \|w - c^{n_j}\| \\
&\qquad\qquad + L\|c^\star - c^{n_j}\| \|F(c^\star) - F(c^{n_j})\| \|w - c^{n_j}\| \\
&\overset{j\to\infty}{\longrightarrow} \quad 0\,.
\end{aligned}
$$

Consequently, for all $w \in B_R$,

$$\langle F'(c^\star)^*(y - F(c^\star)), w - c^\star\rangle \leq 0\,.$$

After the verification of the necessary condition for weak accumulation points we show that the weak convergence of subsequences can be strengthened into convergence in norm topology. This is important to be achieved as in principle our setup is infinite dimensional.

Proposition 6.10. *With the same assumptions as in Proposition 6.9 and the assumptions* (5.6)-(5.7) *on the nonlinear operator F, there exists a subsequence $(c^{n'_l})_{l\in\mathbb{N}} \subset (c^n)_{n\in\mathbb{N}}$ such that $(c^{n'_l})_{l\in\mathbb{N}}$ converges in norm towards the weak accumulation point $c^\star$, i.e.*

$$\lim_{l\to\infty} \|c^{n'_l} - c^\star\| = 0\,.$$

The proof of this proposition is in several parts the same as in [15, Lemma 12]. Here we only mention the difference that is due to the nonlinearity of F. Denote by $(c^{n_j})_{j\in\mathbb{N}}$ the subsequence that was introduced in the proof of Proposition 6.9. Define now $u^j := c^{n_j} - c^\star$, $v^j := c^{n_j+1} - c^\star$, and $\beta^j := \beta^{n_j}$. Due to Lemma 6.8, we have

$\lim_{j\to\infty} \|u^j - v^j\| = 0$. But we also have,

$$
\begin{aligned}
u^j - v^j &= u^j + c^\star - \mathrm{P}_{B_R}(u^j + c^\star + \beta^j F'(v^j + c^\star)^*(y - F(u^j + c^\star))) \\
&= u^j + \mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star))) \\
&\qquad -\mathrm{P}_{B_R}(u^j + c^\star + \beta^j F'(v^j + c^\star)^*(y - F(u^j + c^\star))) \\
&= u^j + \mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star))) \\
&\qquad -\mathrm{P}_{B_R}(c^\star + \beta^j F'(v^j + c^\star)^*(y - F(u^j + c^\star)) + u^j) \qquad (6.10)\\
&\qquad +\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star)) + u^j) \qquad (6.11)\\
&\qquad -\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star)) + u^j) \\
&\qquad +\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(u^j + c^\star)) + u^j) \qquad (6.12)\\
&\qquad -\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(u^j + c^\star)) + u^j)\ , \qquad (6.13)
\end{aligned}
$$

where we have applied Proposition 6.9 ($c^\star$ fulfills the necessary condition) and Lemma 6.3, i.e. $c^\star = \mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star)))$. We consider now the sum of the terms (6.11)+(6.13), and obtain by the assumptions on F and since the β^j are uniformly bounded,

$$
\begin{aligned}
&\|\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star)) + u^j) - \\
&\qquad\qquad \mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(u^j + c^\star)) + u^j)\| \\
&\quad \le \|\beta^j F'(c^\star)^*(F(u^j + c^\star) - F(c^\star))\| \\
&\quad \le \bar\beta \sup_{x\in B_R} \|F'(x)\| \|F(u^j + c^\star) - F(c^\star)\| \overset{j\to\infty}{\longrightarrow} 0\ .
\end{aligned}
$$

The sum of the terms (6.10)+(6.12) yields

$$
\begin{aligned}
&\|\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(u^j + c^\star)) + u^j) - \\
&\qquad\qquad \mathrm{P}_{B_R}(c^\star + \beta^j F'(v^j + c^\star)^*(y - F(u^j + c^\star)) + u^j)\| \\
&\quad \le \bar\beta \left\{ \|(F'(c^\star)^* - F'(v^j + c^\star)^*)(y - F(c^\star))\| \right. \\
&\qquad \left. + \|(F'(c^\star)^* - F'(v^j + c^\star)^*)(F(c^\star) - F(u^j + c^\star))\| \right\} \\
&\quad \le \bar\beta \left\{ \|(F'(c^\star)^* - F'(v^j + c^\star)^*)(y - F(c^\star))\| \right. \\
&\qquad \left. + L\|v^j\| \|F(c^\star) - F(u^j + c^\star)\| \right\} \overset{j\to\infty}{\longrightarrow} 0\ .
\end{aligned}
$$

Consequently, combining $\|u^j - v^j\| \overset{j\to\infty}{\longrightarrow} 0$ and the two last statements, we observe that

$$
\lim_{j\to\infty} \|\mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star)) + u^j) - \mathrm{P}_{B_R}(c^\star + \beta^j F'(c^\star)^*(y - F(c^\star))) - u^j\| = 0\ .
$$

The remaining arguments that verify the strong convergence towards zero of a subsequence of u^j are now the same as in [15, Lemma 12].

As mentioned in [15], one can prove at the cost of more technicalities that the whole subsequence $(c^{n_j})_{j\in\mathbb{N}}$ converges in norm towards $c^\star$. We summarize the findings in the following proposition.

Proposition 6.11. *Every weak accumulation point $c^\star$ of the sequence $(c^n)_{n\in\mathbb{N}}$ defined by* (6.4) *fulfills the necessary condition for a minimizer of $\boldsymbol{D}$ in B_R. Moreover, there exists a subsequence $(c^{n_j})_{j\in\mathbb{N}} \subset (c^n)_{n\in\mathbb{N}}$ that converges in norm to $c^\star$.*

In the next proposition we give a condition under which norm convergence of subsequences carries over to the full sequence $(c^n)_{n\in\mathbb{N}}$.

Proposition 6.12. *Assume that there exists at least one isolated limit $c^\star$ of a subsequence $(c^{n_j})_{j\in\mathbb{N}} \subset (c^n)_{n\in\mathbb{N}}$. Then $c^n \to c^\star$ holds.*

A proof of this assertion can be directly taken from [45].

6.3 Some Algorithmic Aspects

In the previous subsection we have shown norm convergence for all β^n satisfying Condition (B). This, of course, implies also norm convergence for $\beta^n = 1$ for all $n \in \mathbb{N}$, which corresponds to the projected classical Landweber iteration. However, to accelerate the speed of convergence, we are interested in choosing, adaptively, larger values for β^n. In particular, by the reasoning made after Definition 6.5, we like to choose β^n as large as possible. The problem (even for linear operators A) is that the definition of c^{n+1} involves β^n and the inequality (B2) to restrict the choice of β^n uses c^{n+1}. This "implicit" quality does not allow for a straightforward determination of β^n.

For linear problems, conditions (B1) and (B2) are inspired by classical length-step in the steepest descent algorithm for the unconstrained functional $\|Ax-y\|^2$ leading to an accelerated Landweber iteration $x^{n+1} = x^n + \gamma^n A^*(y - Ax^n)$, for which γ^n is picked so that it gives a maximal decrease of $\|Ax-y\|^2$, i.e.

$$\gamma^n = \|A^*(y-Ax^n)\|^2 \|AA^*(y-Ax^n)\|^{-2} .$$

For nonlinear operators this condition translates into a rather non-practical suggestion for γ^n. In our situation, in which we have to fulfill Condition (B), we may derive a much simpler procedure to find a suitable γ^n (which is in our case β^n/r). Due to Lemma 6.8 we have monotonicity of $\boldsymbol{D}$ with respect to the iterates, i.e.

$$L\sqrt{\boldsymbol{D}(c^n)} \le L\sqrt{\boldsymbol{D}(c^{n-1})} \le \ldots \le \frac{r}{2} = \max\{\sup_{c\in B_R} \|F'(c)\|^2, L\sqrt{\boldsymbol{D}(c^0)}\} .$$

Therefore (B3), which was given by

$$L\sqrt{\boldsymbol{D}(c^n)} \leq \beta^n L\sqrt{\boldsymbol{D}(c^n)} \leq \frac{r}{2},$$

is indeed a nontrivial condition for $\beta^n \geq 1$. Namely, the smaller the decrease of $\boldsymbol{D}$, the larger we may choose β^n (when only considering (B3)). Condition (B3) can be recast as $1 \leq \beta^n \leq r/(2L\sqrt{\boldsymbol{D}(c^n)})$ and consequently, by Definition (6.5), an *explicit* (but somewhat "greedy") guess for β^n is given by

$$\beta^n = \max\left\{\sup_{x \in B_R} \frac{\|F'(x)\|^2}{L\sqrt{\boldsymbol{D}(c^n)}}, \sqrt{\frac{\boldsymbol{D}(c^0)}{\boldsymbol{D}(c^n)}}\right\} \geq 1 \, . \tag{6.14}$$

If this choice fulfills (B2) as well, it is retained; if it does not, it can be gradually decreased (by multiplying it with a factor slightly smaller than 1 until (B2) is satisfied.

As a summary of the above reasoning we suggest the following implementation of the proposed projected steepest descent algorithm.

Projected Steepest Descent Method for nonlinear inverse problems	
Given	operator F, its derivative $F'(c)$, data y, some initial guess c^0, and R (sparsity constraint ℓ_1-ball B_R)
Initialization	$r = \max\{2\sup_{c \in B_R} \|F'(c)\|^2, 2L\sqrt{\boldsymbol{D}(c^0)}\}$, set $q = 0.9$ (as an example)
Iteration	**for** $n = 0, 1, 2, \ldots$ **until** a preassigned precision / maximum number of iterations **1.** $\beta^n = \max\left\{\sup_{c \in B_R} \frac{\|F'(c)\|^2}{L\sqrt{\boldsymbol{D}(c^n)}}, \sqrt{\frac{\boldsymbol{D}(c^0)}{\boldsymbol{D}(c^n)}}\right\}$ **2.** $c^{n+1} = \mathrm{P}_{B_R}\left(c^n + \frac{\beta^n}{r} F'(c^{n+1})^*(y - F(c^n))\right)$; by fixed point iteration **3.** verify (B2): $\beta^n \|F(c^{n+1}) - F(c^n)\|^2 \leq \frac{r}{2}\|c^{n+1} - c^n\|^2$ **if** (B2) is satisfied increase n and go to **1.** **otherwise** set $\beta^n = q \cdot \beta^n$ and go to **2.** **end**

6.4 Numerical Experiment: A Nonlinear Sensing Problem

The numerical experiment centers around a nonlinear sampling problem that is very closely related to the sensing problem considered in [54]. The authors of [54] have studied a sensing setup in which a continuous-time signal is mapped by a memoryless, invertible and nonlinear transformation, and then sampled in a non-ideal manner. In this context, memoryless means a static mapping that individually acts at each time instance (pointwise behavior). Such scenarios may appear in acquisition systems where the sensor introduces static nonlinearities, before the signal is sampled by a usual analog-to-digital converter. In [54] a theory and an algorithm is developed that allow a perfect recovery of a signal within a subspace from its nonlinear and non-ideal samples. In our setup we drop the invertibility requirement of the nonlinear transformation, which is indeed quite restrictive. Moreover, we focus on a subclass of problems in which the signal to be recovered is supposed to have sparse expansion.

Let us specify the sensing model. Assume we are given a reconstruction space $\mathcal{A} \subset X$ (e.g. $L_2(\mathbb{R})$) which is spanned by the frame $\{\phi_\lambda : \lambda \in \Lambda\}$ with frame bounds $0 < C_1 \leq C_2 < \infty$. With this frame we associate two mappings, the analysis and synthesis operator,

$$\mathcal{F} : \mathcal{A} \ni f \mapsto \{\langle f, \phi_\lambda \rangle\}_{\lambda \in \Lambda} \in \ell_2(\Lambda) \text{ and } \mathcal{F}^* : \ell_2(\Lambda) \ni x \mapsto \sum_{\lambda \in \Lambda} x_\lambda \phi_\lambda \, .$$

We assume that the function/signal f we wish to recover has a sparse expansion in $\mathcal{A}$. The sensing model is now determined by the nonlinear transformation $M : \mathcal{A} \to Y$ of the continuous-time function f that is point-wise given by the regularized modulus function (to have some concrete example for the nonlinear transformation)

$$M : f \mapsto M(f) = |f|_\varepsilon := \sqrt{f^2 + \varepsilon^2} \, .$$

This nonlinearly transformed f is then sampled in a possibly non-ideal fashion by some sampling function s yielding the following sequence of samples,

$$SM(f) = \{\langle s(\cdot - nT), M(f) \rangle_Y\}_{n \in \mathbb{Z}},$$

where we assume that the family $\{s(\cdot - nT_s), n \in \mathbb{Z}\}$ forms a frame with bounds $0 < S_1 \leq S_2 < \infty$. The goal is to reconstruct f from its samples $y = (S \circ M)(f)$. Since f belongs to $\mathcal{A}$, the reconstruction of f is equivalent with finding a sequence c such that $y = (S \circ M \circ \mathcal{F}^*)(c)$. As $\{\phi_\lambda : \lambda \in \Lambda\}$ forms a frame there might be several different sequences leading to the same function f. Among all possible solutions, we aim (as mentioned above) to find those sequences that have small ℓ_1 norm. As y might be not directly accessible (due to the presence of measurement noise) and due to the nonlinearity of the operator M, it seems more practical not to solve $y = (S \circ M \circ \mathcal{F}^*)(c)$ directly, but to find an approximation $\hat{c}$ such that

$$\hat{c} = \arg\min_{c \in B_R} \|F(c) - y\|^2 \quad \text{and} \, ,$$

where we have used the shorthand notation $F := S \circ M \circ \mathcal{F}^*$ and where the ℓ_1 ball B_R restricts c to have a certain preassigned sparsity.

In order to apply our proposed accelerated steepest descent iteration,

$$c^{n+1} = \mathrm{P}_{B_R}\left(c^n + \frac{\beta^n}{r} F'(c^{n+1})^*(y - F(c^n))\right),$$

to derive an approximation to $\hat{x}$, we have to determine the constants r, see (6.5), and the Lipschitz constant L. This requires a specification of two underlying frames (the reconstruction and sampling frame). One technically motivated choice in signal sampling is the cardinal sine function. This function can be defined as the inverse Fourier transform of the characteristic function of the frequency interval $[-\pi, \pi]$, i.e.

$$\sqrt{2\pi}\operatorname{sinc}(\pi t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \chi_{[-\pi,\pi]}(\omega) e^{it\omega} d\omega \ .$$

Therefore, the resulting function spaces are spaces of bandlimited functions. The inverse Fourier transform of the L_2 normalized characteristic function $\frac{1}{\sqrt{2\Omega}}\chi_{[-\Omega,\Omega]}$ yields

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\Omega}} \chi_{[-\Omega,\Omega]}(\omega) e^{it\omega} d\omega = \sqrt{\frac{\Omega}{\pi}} \operatorname{sinc}(\Omega t)$$

leading to the following definition of L_2 normalized and translated cardinal sine functions,

$$\phi_n(t) = \frac{1}{\sqrt{D_a}} \operatorname{sinc}\left(\frac{\pi}{D_a}(t - nT_a)\right) \ , \quad \text{i.e. } \Omega = \frac{\pi}{D_a} \quad \text{and} \tag{6.15}$$

$$s_n(t) = \frac{1}{\sqrt{D_s}} \operatorname{sinc}\left(\frac{\pi}{D_s}(t - nT_s)\right) \ , \quad \text{i.e. } \Omega = \frac{\pi}{D_s} \tag{6.16}$$

that determine the two frames. The parameters D_a and D_s are fixed and specify here the frequency cut off, whereas T_a and T_s fix the time step sizes. For all $n \in \mathbb{Z}$ we have $\|\phi_n\|_2 = \|s_n\|_2 = 1$. Moreover, it can be easily retrieved that

$$\langle \phi_n, \phi_m \rangle = \operatorname{sinc}\left(\frac{\pi}{D_a}(n-m)T_a\right) \ \text{ and } \ \langle s_n, s_m \rangle = \operatorname{sinc}\left(\frac{\pi}{D_s}(n-m)T_s\right) \ . \tag{6.17}$$

As long as $T_a/D_a, T_s/D_s \in \mathbb{Z}$, the frames form orthonormal systems. The inner products (6.17) are the entries of the Gramian matrices $\mathcal{F}\mathcal{F}^*$ and SS^*, respectively, for which we have $\|\mathcal{F}\mathcal{F}^*\| = \|\mathcal{F}\|^2 = \|\mathcal{F}^*\|^2 \leq C_2$ and $\|SS^*\| = \|S\|^2 = \|S^*\|^2 \leq S_2$.

Let us now determine r and L. To this end we have to estimate $\sup_{c \in B_R} \|F'(c)\|^2$. For given $c \in B_R$, it follows that

$$\begin{aligned} \|F'(c)\| &= \sup_{h \in \ell_2, \|h\|=1} \|F'(c)h\| = \|SM'(\mathcal{F}^*c)\mathcal{F}^*h\| \\ &\leq \|S\| \, \|M'(\mathcal{F}^*c)\| \, \|\mathcal{F}^*\| \ . \end{aligned}$$

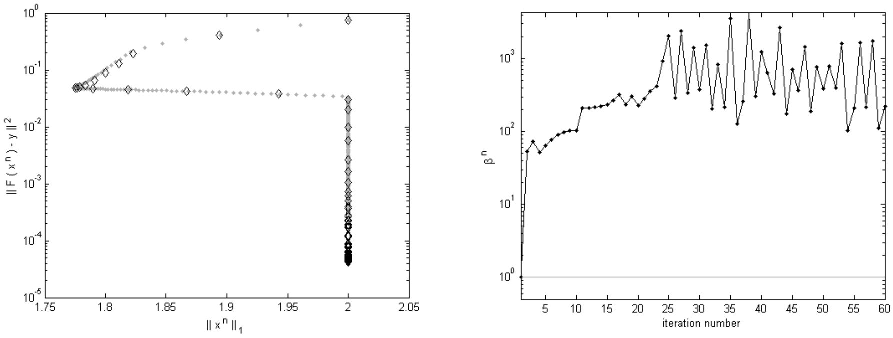

Figure 7. The left image shows the sparsity to residual plot. The black diamonds correspond to the accelerated iteration. For the non-accelerated iteration we have plotted every 20th iteration (gray dots). The right image visualizes the sequence of β^n (black) for the accelerated iteration. The gray line corresponds to $\beta = 1$.

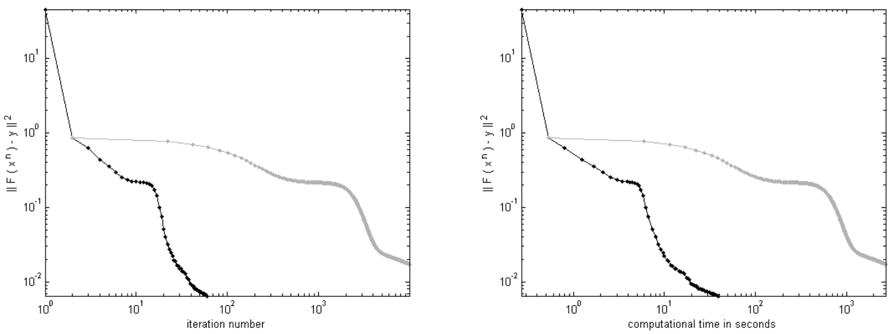

Figure 8. These images represent the residual evolution with respect to the number of iterations (left) and the computational time (right). The black dotted curves represent the residual evolution for the accelerated and the gray dotted curves for the non-accelerated scheme.

Moreover, due to (6.15),

$$
\begin{aligned}
\|M'(\mathcal{F}^*c)\|^2 &= \sup_{h\in\Lambda_2,\|h\|=1} \|M'(\mathcal{F}^*c)h\|^2 \\
&= \int_{\mathbb{R}} |(\mathcal{F}^*c)(t)|^2|((\mathcal{F}^*c)(t))^2+\varepsilon^2|^{-1}|h(t)|^2dt \\
&\leq \frac{1}{\varepsilon^2}\int_{\mathbb{R}}(\sum_n |c_n||\phi_n(t)|)^2|h(t)|^2dt \leq \frac{\|c\|_1^2}{\varepsilon^2 D_a} .
\end{aligned}
$$

Therefore, we finally obtain

$$\sup_{c\in B_R} \|F'(c)\|^2 \leq \|S\|^2\,\|\mathcal{F}^*\|^2\,\frac{R^2}{\varepsilon^2 D_a} \leq S_2\,C_2\,\frac{R^2}{\varepsilon^2 D_a}\,. \tag{6.18}$$

The Lipschitz continuity of F' is characterized by $\|F'(\tilde{c}) - F'(c)\| \leq L\|\tilde{c} - c\|$, for all $c, \tilde{c} \in B_R$. In order to find the Lipschitz constant L, we directly derive

$$\begin{aligned}
\|F'(\tilde{c}) - F'(c)\| &= \sup_{h\in\ell_2, \|h\|=1} \|F'(\tilde{c})h - F'(c)h\| \\
&= \sup_{h\in\ell_2, \|h\|=1} \|SM'(\mathcal{F}^*\tilde{c})\mathcal{F}^*h - SM'(\mathcal{F}^*c)\mathcal{F}^*h\| \\
&\leq \|S\|\,\|M'(\mathcal{F}^*\tilde{c}) - M'(\mathcal{F}^*c)\|\,\|\mathcal{F}^*\|\,, \qquad (6.19)
\end{aligned}$$

and with $M''(f) = \varepsilon^2(f^2 + \varepsilon^2)^{-3/2}$ it follows

$$\begin{aligned}
&\|M'(\mathcal{F}^*\tilde{c}) - M'(\mathcal{F}^*c)\|^2 \\
&\quad= \sup_{h\in L_2, \|h\|=1} \int_{\mathbb{R}} |M'(\mathcal{F}^*\tilde{c}(t)) - M'(\mathcal{F}^*c(t))|^2 |h(t)|^2 dt \\
&\quad\leq \sup_{h\in L_2, \|h\|=1} \int_{\mathbb{R}} \frac{1}{\varepsilon^2} |\mathcal{F}^*\tilde{c}_n(t) - \mathcal{F}^*c(t)|^2\,|h(t)|^2 dt \\
&\quad\leq \sup_{h\in L_2, \|h\|=1} \int_{\mathbb{R}} \frac{1}{\varepsilon^2} \left(\sum_{n\in\mathbb{Z}} |(\tilde{c}_n - c_n)|\,|\phi_n(t)|\right)^2 |h(t)|^2 dt \\
&\quad\leq \sup_{h\in L_2, \|h\|=1} \int_{\mathbb{R}} \sum_{n\in\mathbb{Z}} |\phi_n(t)|^2\,|h(t)|^2 dt\,\frac{1}{\varepsilon^2}\|\tilde{c} - c\|^2\,.
\end{aligned}$$

To finally bound the last quantity, we have to estimate $\sum_{n\in\mathbb{Z}} |\phi_n(t)|^2$ independently on $t \in \mathbb{R}$. With definition (6.15), we observe that

$$\sum_{n\in\mathbb{Z}} |\phi_n(t)|^2 = \frac{1}{D_a} \sum_{n\in\mathbb{Z}} \operatorname{sinc}^2\left(\frac{\pi}{D_a}t - n\frac{\pi\,T_a}{D_a}\right) \tag{6.20}$$

is a periodic function with period T_a. Therefore it is sufficient to analyze (6.20) for $t \in [0, T_a]$. The sum in (6.20) is maximal for $t = 0$ and $t = T_a$. Consequently, with

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} \operatorname{sinc}^2\left(n\frac{\pi\,T_a}{D_a}\right) &= 1 + \sum_{n\in\mathbb{Z}\setminus\{0\}} \operatorname{sinc}^2\left(n\frac{\pi\,T_a}{D_a}\right) \leq 1 + \frac{2\,D_a^2}{\pi^2\,T_a^2} \sum_{n\in\mathbb{N}\setminus\{0\}} \frac{1}{n^2} \\
&= 1 + \frac{4\,D_a^2}{\pi^2\,T_a^2}
\end{aligned}$$

we obtain by combining (6.19) and (6.20),

$$\|F'(\tilde{c}) - F'(c)\| \leq L\,\|\tilde{c} - c\|\ , \quad \text{with } L := \frac{1}{\varepsilon}\sqrt{\frac{1}{D_a} + \frac{4\,D_a}{\pi^2\,T_a^2}}\sqrt{S_2}\,\sqrt{A_2}\ . \tag{6.21}$$

In our concrete example (visualized in Figure 9) the ansatz space $\mathcal{A} \subset L_2(\mathbb{R})$ is spanned by functions a_n with $D_a = 0.4$ and time step size $T_a = 0.1$. The sampling map S is determined by $D_s = 0.2$ and $T_s = 0.1$. The synthetic signal which we aim to reconstruct is given by

$$f(t) = a_{-2}(t) - 0.5a_{2.5}(t)\ .$$

For the numerical implementation we have restricted the computations to the finite interval $[-10, 10]$ which was discretized by the grid $t_k = -10 + 0.05\,k$ with $k = 0, 1, 2, \ldots$. The bounds A_2 and S_2 are estimated by the eigenvalues of adequately corresponding finite dimensional approximations of the Gramian matrices $\langle a_n, a_m\rangle$ and $\langle s_n, s_m\rangle$. For the radius of the ℓ_1 ball (determined the sparsity constraint) we have picked $R = 2$. This choice of course includes some a-priori knowledge of the solution to be reconstructed. Usually there is no a-priori information on R available. Even if not proven so far, R plays the role of an regularization parameter (so far just with numerical evidence). Therefore, we can observe in case of misspecified R a similar behavior as for inversion methods where the regularization parameter was not optimally chosen. If R is chosen too large it may easily happen that the ℓ_1 constraint has almost no impact and the solution can be arbitrarily far off the true solution. Therefore, it was suggested in [15] to choose a slowly increasing radius, i.e.

$$R^n = (n+1)R/N\ ,$$

where n is the iteration index and N stands for a prescribed number of iterations. This proceeding yields in all considered experiments better results. However, convergence of a scheme with varying R^n is theoretically not verified yet.

In Figure 7 (right image) one finds that β^n varies significantly from one to another iteration. This verifies the usefulness of Condition (B). From the first iteration on, the values for β^n are obviously larger than one and grow in the first phase of the iteration process (for the accelerated method only the first 60 iterations are shown). But the main impact manifests itself more in the second half of the iteration ($n > 20$) where the non-accelerated variant has a much less decay of $\sqrt{\boldsymbol{D}(x^n)}$, see Figure 8. There the values of β^n vary around 10^3 and allow that impressive fast and rapid decay of $\sqrt{\boldsymbol{D}(x^n)}$ of the accelerated descent method. For the non-accelerated method we had to compute 10^4 iterations to achieve reasonable small residuals $\sqrt{\boldsymbol{D}(x^n)}$ (but even then being far off the nice results achieved by the accelerated scheme). The right plot in Figure 8 sketches the residual decay with respect to the overall computational time that was practically necessary. Both curves (the black and the gray)

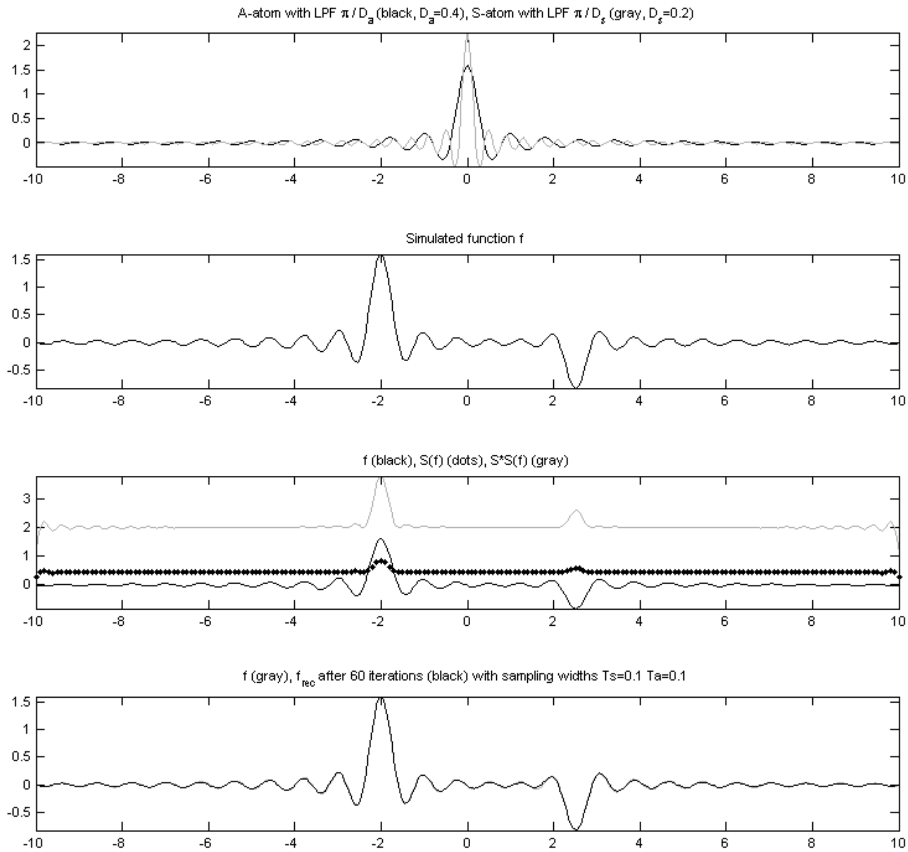

Figure 9. This overview plot shows the used atoms a_0 and s_0 (1st row), the simulated signal (2nd row), the nonlinearly and non-ideally sampled values (3rd row), and the final approximation $A^*x^{60} \in \mathcal{A}$ that was computed with accelerated iteration scheme.

were of course obtained on the same machine under same conditions. The achieved time reduction is remarkable as the accelerated iteration method has required many additional loops of the individual fixed point iterations in order to find the optimal β^n. In particular, the final residual value after $n = 10.000$ iterations for the non-accelerated method was $\sqrt{\boldsymbol{D}(x^{10000})} = 0.0172$. This value was reached by the accelerated method after $n = 28$ iteration steps (the final value after $n = 60$ iterations was $\sqrt{\boldsymbol{D}(x^{60})} = 0.0065$). The overall computational time consumption of the non-accelerated method to arrive at $\sqrt{\boldsymbol{D}(x^{10000})} = 0.0172$ was 45min and 2s, whereas the time consumption for the accelerated method for the same residual discrepancy was only 11.8s, i.e. 229 times faster. The finally resulting reconstruction including a diagram showing the nonlinearly sampled data is given in Figure 9.

Summarizing this numerical experiment, we can conclude that all the theoretical statements of the previous sections can be verified. For this particular nonlinear sensing problem we can achieve an impressive factor of acceleration. But this, however, holds for this concrete setting. There is no proved guaranty that the same can be achieved for other applications.

Acknowledgments. G.T. and R.R. would like to thank M. Fornasier for the kind invitation to contribute to the present book. G.T. especially acknowledges the vivid and fruitful interaction with the research team of R. Ramlau and the very nice and close collaboration within his own research group (thanks C. Borries, E. Herrholz, M. Nitschke, K. Pönitz, M. Rätsch, C. Reitberger). R.R. thanks in particular M. Zhariy, S. Anzengruber and E. Resmerita for their collaboration and support in producing this contribution. Doing joint work with G.T. was always enjoyable.
The research of G.T. was partially supported by Deutsche Forschungsgemeinschaft Grants TE 354/3-1, TE 354/4-1, TE 354/5-1, TE 354/8-1. The research of R.R. was supported by the Austrian Science Fund, Grants P20237-N14, P19496-N18 and DK W1214-N15

Bibliography

[1] S. Anzengruber and R. Ramlau, Morozovs discrepancy principle for Tikhonov-type functionals with non-linear operators. *Inverse Problems* 26 (2), (2010), 1–17.

[2] K. Bredies, D.A. Lorenz and P. Maass, A Generalized Conditional Gradient Method and its Connection to an Iterative Shrinkage Method, *Computational Optimization and Application* 42 (2009), 173–193.

[3] E. J. Candès and F. Guo, New Multiscale Transforms, Minimum Total Variation Synthesis: Application to Edge-Preserving Image Restoration, *Preprint CalTech* (2001).

[4] Y. Censor, D. Gustafson, A. Lent and H. Tuy, A new approach to the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients, *IEEE Trans. Nucl. Sci.* (1979), 2275–79.

[5] A. CHAMBOLLE, R. DEVORE, N. Y. LEE AND B. LUCIER, Non-linear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage, *IEEE Tran. Image Proc.*, 7 (1998), 319–333.

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

[7] P.G. Ciarlet, *Introduction to Numerical Linear Algebra and Optimisation*, Cambrigde Univ. Pr., Cambridge, 1995.

[8] A. Cohen, *Numerical Analysis of wavelet methods*, 32, North-Holland, Amsterdam, 2003.

[9] A. Cohen, W. Dahmen, I. Daubechies and R. DeVore, *Harmonic Analysis of the Space BV*, IGPM Report # 195, RWTH Aachen, 2000.

[10] A. Cohen, W. Dahmen and R. DeVore, Adaptive wavelet methods II - Beyond the elliptic case, *Found. Comput. Math.* 2 (2002), 203–245.

[11] A. Cohen, R. DeVore, P. Petrushev and H. Xu, Nonlinear Approximation and the Space $BV(\mathbb{R}^2)$, *American Journal of Mathematics* (1999), 587–628.

[12] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[13] I. Daubechies, M. Defrise and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.* 51 (2004), 1413–1541.

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

[15] 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).

[16] 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.* (2004).

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

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

[19] ______, On some iterative concepts for image restoration, *Advances in Imaging and Electron Physics* 150 (2008), 1–51.

[20] M. Defrise and C. De Mol, *A note on stopping rules for iterative methods and filtered SVD*, Inverse Problems: An Interdisciplinary Study (P. C. Sabatier, ed.), Academic Press, 1987, pp. 261–268.

[21] V. Dicken, *Simultaneous Activity and Attenuation Reconstruction in Single Photon Emission Computed Tomography, a Nonlinear Ill-Posed Problem*, Ph.d. thesis, Universität Potsdam, 5/ 1998.

[22] ______, A new approach towards simultaneous activity and attenuation reconstruction in emission tomography, *Inverse Problems* 15 (1999), 931–960.

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

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

[25] M. Fornasier and H. Rauhut, Recovery Algorithms for Vector Valued Data with Joint Sparsity Constraint, *Preprint* (2006).

[26] M. GRASSMAIR, M. HALTMEIER AND O. SCHERZER, Sparse regularization with ℓ_q penalty term, *Inverse Problems*, 24 (2008), pp. 1–13.

[27] E. Klann, P. Maass and R. Ramlau, Two-step regularization methods for linear inverse problems, *J. Inverse Ill-Posed Probl.*, 14 (2006), pp. 583–609.

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

[29] S. H. Manglos and T. M. Young, *Constrained IntraSPECT Reconstructions from SPECT Projections*, Conf. Rec. IEEE Nuclear Science Symp. and Medical Imaging Conference, San Francisco, CA, 1993, pp. 1605–1609.

[30] P. Mathé and S. Pereverzev, Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods, *SIAM J. Numer. Anal.*, 38 (2001), pp. 1999–2021.

[31] A. Neubauer, When do Sobolev spaces form a Hilbert scale , *Proc. Am. Math. Soc.*, 103 (1988), 557–562.

[32] R.D. Nowak M.A.T. Figueiredo and S.J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE J. Sel. Top. Signal Process.* 1 (2007).

[33] Y. Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, *University Lecture Series Volume 22, AMS* (2002).

[34] ———, Oscillating Patterns in some Nonlinear Evolution Equations, *CIME report* (2003).

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

[36] S. Osher and L. Vese, *Modeling textures with total variation minimization and oscillating patterns in image processing*, University of California Los Angeles C.A.M., Report no. 02-19, 2002.

[37] R. Ramlau, *Modifizierte Landweber-Iterationen für Inverse Probleme*, PhD thesis, Universität Potsdam (1997).

[38] ———, A modified Landweber–Method for Inverse Problems, *Journal for Numerical Functional Analysis and Optimization* 20 (1999), 79–98.

[39] R. Ramlau, Morozov's Discrepancy Principle for Tikhonov regularization of nonlinear operators, *Numer. Funct. Anal. and Optimiz* 23 (2002), 147–172.

[40] ———, A steepest descent algorithm for the global minimization of the Tikhonov– functional, *Inverse Problems* 18 (2002), 381–405.

[41] ———, TIGRA–an iterative algorithm for regularizing nonlinear ill–posed problems, *Inverse Problems* 19 (2003), 433–467.

[42] ———, On the use of fixed point iterations for the regularization of nonlinear ill-posed problems, *Journal for Inverse and Ill-Posed Problems* 13 (2) (2005),175–200.

[43] R. Ramlau, R. Clackdoyle, F. Noo and G. Bal, Accurate attenuation correction in SPECT imaging using optimization of bilinear functions and assuming an unknown spatially–varying attenuation distribution, *Z. Angew. Math. Mech.* 80 (2000), 613–621.

[44] R. Ramlau and G. Teschke, Tikhonov Replacement Functionals for Iteratively Solving Nonlinear Operator Equations, *Inverse Problems* 21 (2005), 1571–1592.

[45] ______, Tikhonov Replacement Functionals for Iteratively Solving Nonlinear Operator Equations, *Inverse Problems* 21 (2005), 1571–1592.

[46] ______, A Projection Iteration for Nonlinear Operator Equations with Sparsity Constraints, *Numerische Mathematik* 104 (2006), 177–203.

[47] R. Ramlau, G. Teschke and M. Zhariy, Nonlinear and Adaptive Frame Approximations Schemes for Ill-Posed Inverse Problems: Regularization Theory and Numerical Experiments. *Inverse Problems* 24 (2008) 065013.

[48] R. Ramlau and E. Resmerita, Convergence rates for regularization with sparsity constraints, *ETNA* 37 (2010), 87 – 104.

[49] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variations based noise removal algorithms, *Physica D* 60 (1992), 259–268.

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

[51] J. A. Terry, B. M. W. Tsui, J. R. Perry, J. L. Hendricks and G. T. Gullberg, *The design of a mathematical phantom of the upper human torso for use in 3-D SPECT imaging research*, Proc. 1990 Fall Meeting Biomed. Eng. Soc. (Blacksburg, VA), New York University Press, 1990, pp. 1467–74.

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

[53] G. Teschke and C. Borries, Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints, *Inverse Problems* 26 (2010), 025007.

[54] Y. Eldar T.G. Dvorkind and E. Matusiak, Nonlinear and Non-Ideal Sampling: Theory and Methods, *IEEE Trans. on Signal Processing* 56 (2008).

[55] E. van den Berg and M.P. Friedlander, In pursiut of a root, *Preprint* (2007).

[56] A. Welch, R. Clack, P. E. Christian and G. T. Gullberg, Toward accurate attenuation correction without transmission measurements, *Journal of Nuclear Medicine* (1996), 18P.

[57] A Welch, R Clack, F Natterer and G T Gullberg, Toward accurate attenuation correction in SPECT without transmission measurements, *IEEE Trans. Med. Imaging* (1997), 532–40.

[58] E. Zeidler, *Nonlinear Functional Analysis and its Applications III*, Springer, New York, 1985.

[59] ______, *Nonlinear Functional Analysis and its Applications I*, Springer, New York, 1986.

Author information

Ronny Ramlau, Industrial Mathematics Institute, Johannes Kepler University Linz,
Altenbergerstraße 69, A-4040 Linz, Austria.
E-mail: `ronny.ramlau@jku.at`

Gerd Teschke, Institute for Computational Mathematics in Science and Technology, Neubrandenburg University of Applied Sciences, Brodaer Str. 2, 17033 Neubrandenburg, Germany.
E-mail: teschke@hs-nb.de