Nr. 64
27. September 2019

universität innsbruck
Institut für Mathematik

Preprint-Series: Department of Mathematics - Applied Mathematics

Sparse regularization of inverse problems by operator-adapted frame thresholding

J. Frikel, M. Haltmeier

REVISED VERSION, DECEMBER 2019

APPLIEDMATHEMATICS

Technikerstraße 13 - 6020 Innsbruck - Austria
Tel.: +43 512 507 53803 Fax: +43 512 507 53898
https://applied-math.uibk.ac.at

Sparse regularization of inverse problems by operator-adapted frame thresholding

Jürgen Frikel

Department of Computer Science and Mathematics, OTH Regensburg
Galgenbergstraße 32, D-93053 Regensburg, Germany
juergen.frikel@oth-regensburg.de

Markus Haltmeier

Department of Mathematics, University of Innsbruck
Technikerstrasse 13, 6020 Innsbruck, Austria
markus.haltmeier@uibk.ac.at

Abstract

We analyze sparse frame based regularization of inverse problems by means of a diagonal frame decomposition (DFD) for the forward operator, which generalizes the SVD. The DFD allows to define a non-iterative (direct) operator-adapted frame thresholding approach which we show to provide a convergent regularization method with linear convergence rates. These results will be compared to the well-known analysis and synthesis variants of sparse ℓ^1-regularization which are usually implemented thorough iterative schemes. If the frame is a basis (non-redundant case), the three versions of sparse regularization, namely synthesis and analysis variants of ℓ^1-regularization as well as the DFD thresholding are equivalent. However, in the redundant case, those three approaches are pairwise different.

1 Introduction

This paper is concerned with inverse problems of the form

$$\boldsymbol{y}^\delta = \mathbf{A}\boldsymbol{x} + \boldsymbol{z}\,, \tag{1.1}$$

where $\mathbf{A}\colon \mathbb{D}(\mathbf{A}) \subseteq \mathbb{X} \to \mathbb{Y}$ is a linear operator between Hilbert spaces, and $\boldsymbol{z}$ denotes the data distortion (noise). We allow unbounded operators and assume that $\mathbb{D}(\mathbf{A})$ is dense. Moreover, we assume that the unknown object $\boldsymbol{x}$ is an element of a closed subspace space $\mathbb{X}_0 \subseteq \mathbb{X}$ on which $\mathbf{A}$ is bounded. We are particularly interested in problems, where (1.1) is ill-posed in which case the solution of (1.1) (if existent) is either not unique or the solution operator is not continuous (hence, the solution process is unstable with respect to data perturbations). In order to stabilize the inversion of (1.1) one has to apply regularization methods, cf. [10, 22]. The basic idea of regularization is to include a-priori information about the unknown object into the solution process.

In this paper, we use sparsity based regularization, where the a-priori assumption on the unknown object is sparsity of x with respect to a frame $(u_\lambda)_{\lambda\in\Lambda}$ of $\mathbb{X}$, cf. [22, 13, 6, 16, 20, 2]. That is, we regularize the recovery of x from measurements (1.1) by enforcing sparsity of x with respect to a suitably chosen frame of $\mathbb{X}$. Sparse regularization is well investigated and has been applied to many different imaging problems, and by now there are many algorithms available that implement sparse regularization. However, when dealing with frames, there are at least two fundamentally different concepts implementing sparsity, namely the synthesis and the analysis variant. The reason for this lies in the fact that expansions of $x \in \mathbb{X}$ with respect to frames are not unique (which is in contrast to basis expansions). In the synthesis variant, it is assumed that the unknown is a sparse linear combination of frame elements, whereas, in the analysis variant, it is required that the inner products $\langle u_\lambda, x\rangle$ with respect to a given frame are sparse. The difference between these approaches has been pointed out clearly in [9].

Sparse regularization is widely used in inverse problems as it provides good regularization results and is able to preserve or emphasize features (e.g., edges) in the reconstruction. However, this often comes at the price of speed, since most of the algorithms implementing sparse regularization are based on variational formulations that are solved by iterative schemes.

In the present paper, we investigate a third variant of sparse regularization that is based on an operator-adapted diagonal frame decomposition (DFD) of the unknown object, cf. [8, 3, 5] and which allows to define a direct (non-iterative) sparse regularization method. In the noise-free case ($z = 0$), explicit reproducing formulas for the unknown object can be derived from the DFD, where the frame coefficients of x are calculated directly from the data $y = \mathbf{A}x$. In the presence of noise ($z \neq 0$), regularized versions of those formulas are obtained by applying component-wise soft-thresholding to the calculated coefficients, where the soft-thresholding operator is defined as follows:

Definition 1.1 (Soft-thresholding). *Let Λ be some index set.*

- *For $\eta, d \in \mathbb{K}$ let* $\operatorname{soft}(\eta, d) := \operatorname{sign}(\eta)\max\{0, |\eta| - d\}$.
- *For $\boldsymbol{\eta}, \boldsymbol{d} \in \mathbb{K}^\Lambda$ we define the component-wise soft-thresholding by*

$$\mathbb{S}_{\boldsymbol{d}}(\boldsymbol{\eta}) := (\operatorname{soft}(\eta_\lambda, d_\lambda))_{\lambda\in\Lambda}\,. \tag{1.2}$$

Here and below we define $\operatorname{sign}(\eta) := \eta/|\eta|$ for $\eta \in \mathbb{K}\setminus\{0\}$ and $\operatorname{sign}(0) := 0$. The advantage of the DFD-variant of sparse regularization lies in the fact that it admits a non-iterative (direct) and fast implementation of sparse regularization that can be easily implemented for several inverse problems.

We point out, that the three variants of sparse regularization (mentioned above) are equivalent if orthonormal bases are used instead of frames, but they are fundamentally different in the redundant case.

As the main theoretical results in this paper we show that the third variant of sparse regularization, which we call *DFD-thresholding*, defines a convergent regularization method and we derive linear convergence rates for sparse solutions. For the basis case, the same results follow from existing results of ℓ^1-regularization [6, 13, 14]. In the redundant case, the results follow from [14] for the synthesis approach and from [15] for the analysis approach. In case of DFD-thresholding, we are not aware of any results concerning convergence analysis or convergence rates.

Outline

This paper is organized as follows. In Section 2 we define the diagonal frame decompositions of operators and give several examples of diagonal frame expansions for various operators using wavelet, curvelet, and shearlet frames. In Section 3 we review the convergence theory of ℓ^1-regularization and the convergence rates. In Section 4, we show that DFD-thresholding is a convergent regularization method and derive its convergence rates.

2 Diagonal frame decomposition

In this section, we introduce the concept of operator adapted diagonal frame decompositions (DFD) and discuss some classical examples of such DFDs in the case of the classical 2D Radon transform and the forward operator of photoacoustic tomography with a flat observation surface.

2.1 Formal definition

We define the operator adapted diagonal frame decomposition as a generalization of the wavelet vaguelette decomposition and the biorthogonal curvelet or shearlet decompositions to general frames, cf. [8, 3, 5].

Definition 2.1 (Diagonal frame decomposition (DFD)). *Let $(u_\lambda)_{\lambda\in\Lambda} \in \mathbb{X}^\Lambda$, $(v_\lambda)_{\lambda\in\Lambda} \in \mathbb{Y}^\Lambda$ and let $(\kappa_\lambda)_{\lambda\in\Lambda}$ be a family of positive numbers. For a linear operator $\mathbf{A}\colon \mathbb{D}(\mathbf{A}) \subseteq \mathbb{X} \to \mathbb{Y}$, we call $(u_\lambda, v_\lambda, \kappa_\lambda)_{\lambda\in\Lambda}$* a diagonal frame decomposition (DFD) for $\mathbf{A}$, *if the following conditions hold:*

- *(D1) $(u_\lambda)_\lambda$ is a frame of $\mathbb{X}$,*
- *(D2) $(v_\lambda)_\lambda$ is a frame of $\overline{\operatorname{ran}(\mathbf{A})} = \overline{\mathbf{A}(\mathbb{X})}$,*
- *(D3) $\forall \lambda \in \Lambda\colon \kappa_\lambda \neq 0 \wedge \mathbf{A}^* v_\lambda = \kappa_\lambda u_\lambda$.*

Remark 2.2. *The DFD generalizes the singular value decomposition (SVD) and the wavelet-vaguelette decomposition (WVD) (cf. [8]) as it allows the systems $(u_\lambda)_\lambda$ and $(v_\lambda)_\lambda$ to be non-orthogonal and redundant. Note that by (D2) and (D3), the frame $(u_\lambda)_\lambda$ satisfies $u_\lambda \in \mathbf{A}^*(\overline{\operatorname{ran}(\mathbf{A})}) = \operatorname{ran}(\mathbf{A}^*)$, where we have made use of the identity $\operatorname{ran}(\mathbf{A})^\perp = \ker(\mathbf{A}^*)$. For typical inverse problems this yields a notable smoothness assumption on the involved elements of the frame $(u_\lambda)_\lambda$.*

Although the SVD has proven itself to be a useful tool for analyzing and solving inverse problems it has the following drawbacks: First, ONBs that are provided by the SVD (though optimally adapted to the operator in consideration), in many cases, don't provide sparse representations of signals of interest and, hence, are not suitable for the use in sparse regularization. In particular, frames that provide sparse representation of signals (such as wavelets or wavelet-like systems) are often not part of the SVD. Second, SVD is often very hard to compute and not known analytically for many practical applications.

To overcome some of those difficulties, wavelet-vaguelette decompositions were introduced. Nevertheless, this concept builds upon expansions of signals with respect to orthogonal wavelet-systems, which may not provide an optimal sparse representation of signals of interest, e.g., signals with sharp edges. Thus, by allowing general frames, the BCD offers great flexibility in the choice of a suitable function system for sparse regularization while retaining the advantages.

Definition 2.3. *For a frame $(u_\lambda)_{\lambda\in\Lambda}$ of $\mathbb{X}$, the* synthesis operator *is defined as* $\mathbf{U}\colon \ell^2(\Lambda) \to \mathbb{Y}\colon \boldsymbol{\xi} \mapsto \sum_{\lambda\in\Lambda} \xi_\lambda u_\lambda$ *and the corresponding* analysis operator *is defined as its adjoint,* $\mathbf{U}^*\colon \mathbb{X} \to \ell^2(\Lambda)\colon \boldsymbol{x} \mapsto (\langle \boldsymbol{x}, u_\lambda\rangle)_{\lambda\in\Lambda}$.

In what follows, the synthesis operator of a frame will be always denoted with the corresponding upper case letter, e.g., if $(v_\lambda)_{\lambda\in\Lambda}$ is a frame, then $\mathbf{V}$ denotes the corresponding synthesis and $\mathbf{V}^*$ the analysis operator.

In order to simplify the notation, we will also refer to $(\mathbf{U}, \mathbf{V}, \boldsymbol{\kappa})$ the as DFD instead of using the full notation $(u_\lambda, v_\lambda, \kappa_\lambda)_{\lambda\in\Lambda}$.

If a DFD exists for an operator $\mathbf{A}$, it immediately gives rise to a *reproducing formula*

$$\boldsymbol{x} = \sum_{\lambda\in\Lambda} \langle \boldsymbol{x}, u_\lambda\rangle\, \bar{u}_\lambda = \sum_{\lambda\in\Lambda} \kappa_\lambda^{-1} \langle \mathbf{A}\boldsymbol{x}, v_\lambda\rangle\, \bar{u}_\lambda = \bar{\mathbf{U}} \circ \mathbf{M}_{\boldsymbol{\kappa}}^{+} \circ \mathbf{V}^*(\mathbf{A}\boldsymbol{x}), \tag{2.1}$$

where $(\bar{u}_\lambda)_{\lambda\in\Lambda}$ is the dual frame to $(u_\lambda)_{\lambda\in\Lambda}$ (cf. [4]) and $\bar{\mathbf{U}}$ the corresponding synthesis operator. Moreover, $\mathbf{M}_{\boldsymbol{\kappa}}^{+}$ denotes the Moore-Penrose inverse of $\mathbf{M}_{\boldsymbol{\kappa}}$ and performs pointwise division with $\boldsymbol{\kappa}$, i.e.

$$(\mathbf{M}_{\boldsymbol{\kappa}}^{+}(\boldsymbol{\eta}))_\lambda := \begin{cases} (\eta_\lambda/\kappa_\lambda)_{\lambda\in\Lambda} & \text{if } \kappa_\lambda \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.2}$$

Hence, from given (clean) data $\boldsymbol{y}$, one can calculate the frame coefficient of $\boldsymbol{x}$ and obtain a reconstruction via (2.1). The key to the practical use of this reproducing formulas is the efficient implementation of the analysis and synthesis operators $\mathbf{V}^*$ and $\bar{\mathbf{U}}$, respectively. For particular cases, we will provide efficient and easy to implement algorithms for the evaluation of $\mathbf{V}^*$ and $\bar{\mathbf{U}}$.

Note that, the reproducing formula (2.1) (similarly to the SVD) reveals the ill-posedness of the operator equation through the decay of the *quasi-singular values* κ_λ. A regularized version of the reproducing formula (2.1) can be obtained by incorporating soft-thresholding of the frame coefficients. In section 4, we present a complete analysis of this approach as we are not aware of any results for the general DFD-thresholding in the context of regularization theory.

We now provide several examples of DFDs, including the wavelet vaguelette decomposition and the biorthogonal curvelet decomposition [3] for the Radon transform as well as a DFD for the forward operator of photoacoustic tomography with flat observation surface.

2.2 Radon transform

Definition 2.4 (Radon transform). *The Radon transform* $\mathbf{R}\colon L^2(B_1(0)) \to L^2(\mathbb{S}^1 \times \mathbb{R})$ *is defined by*

$$\forall (\theta, s) \in \mathbb{S}^1 \times \mathbb{R}\colon \quad \mathbf{R}f(\theta, s) = \int_{\mathbb{R}} f(s\theta + t\theta^\perp)\, \mathrm{d}s\,. \tag{2.3}$$

It is well known that the Radon transform is bounded on $L^2(B_1(0))$, see [19]. Let $\mathbf{F}g(\theta,\omega) = \int_{\mathbb{R}} g(\theta,s)e^{-i\omega s}$ be the Fourier transform with respect to the first component and consider the Riesz potential [19]

$$(\mathbf{I}^{-\alpha}g)(\theta,\omega) := \frac{1}{2\pi}\int_{\mathbb{R}} |\omega|^{\alpha}\,(\mathbf{F}g)(\theta,\omega)e^{i\omega s}\,\mathrm{d}s \tag{2.4}$$

for $\alpha > -1$. The following hold:

(R1) The commutation relation $(\mathbf{I}^{-\alpha}\circ\mathbf{R})f = (\mathbf{R}\circ(-\Delta)^{\alpha/2})f$.

(R2) The filtered backprojection formula $f = (4\pi)^{-1}\mathbf{R}^*(\mathbf{I}^{-1}\circ\mathbf{R})f =: \mathbf{R}^\sharp\mathbf{R}f$.

(R3) Isometry property $4\pi\,\langle f_1, f_2\rangle_{L^2} = \langle \mathbf{I}^{-1}\circ\mathbf{R}f_1, \mathbf{R}f_2\rangle_{L^2}$.

Using these ingredients, one can obtain a DFD for the Radon transform as follows:

Example 2.5. *DFD for the Radon transform Let $(u_\lambda)_{\lambda\in\Lambda}$ be either an orthonormal basis of wavelets with compact support, a (band-limited) curvelet or a shearlet tight frame with $\lambda = (j,k,\beta)\in\Lambda$ where $j\geq 0$ is the scale index. Then $(\mathbf{U},\mathbf{V},\boldsymbol{\kappa})$ is a DFD with*

$$v_\lambda := 2^{-j/2}\,(4\pi)^{-1}(\mathbf{I}^{-1}\circ\mathbf{R})u_\lambda \tag{2.5}$$

$$\kappa_\lambda := 2^{-j/2}\,. \tag{2.6}$$

These results have been obtained in [8] for wavelet bases, in [3] for curvelet systems and in [5] for the shearlet frame. All cases are shown in similar manner and basically follow from (R1), (R2) and the fact that that $2^{-j/2}(-\Delta)^{1/4}(4\pi)^{-1}u_\lambda \simeq u_\lambda$ for any of the considered systems. The limited data case has been studied in [11].

Equation (2.5) implies

$$\langle g, v_\lambda\rangle = 2^{-j/2}(4\pi)^{-1}\left\langle g, \mathbf{I}^{-1}\circ\mathbf{R}u_\lambda\right\rangle = 2^{-j/2}(4\pi)^{-1}\left\langle \mathbf{R}^*\circ\mathbf{I}^{-1}g, u_\lambda\right\rangle\,. \tag{2.7}$$

This gives an efficient numerical algorithm for the evaluation of $\mathbf{V}^*$ provided that $\mathbf{U}^*$ is associated with an efficient algorithm. This is in particular the case for the wavelet, shearlet and curvelet frames as above.

Remark 2.6. *We would like to note that in order to define a DFD for the case of curvelets or shearlets one needs to consider the Radon transform on subspaces of $L^2(\mathbb{R}^2)$ consisting of functions that are defined on unbounded domains (since band-limited curvelets or shearlets have non-compact support). However, since the Radon transform is an unbounded operator on $L^2(\mathbb{R}^2)$, the reproducing formula* (2.1) *will not hold any more in general. The reproducing formulas are at least available for the case that the object x can be represented as a finite linear combination of curvelets or shearlets (cf. [3] and [5]).*

Another possibility would be to consider projections of curvelet or shearlet frames onto the space $L^2(B_1(0))$, which would yield a frame for this space (cf. [4]) and then define the DFD in the same way as above. Because the Radon transform is continuous on $L^2(B_1(0))$, the reproducing formula (2.1) *will hold for general linear combinations.*

Algorithm 2.7. *Computing DFD coefficients for the Radon transform Let* $\mathbf{U}$ *be a wavelet, shearlet of curvelet frame and define* $\mathbf{V}$ *by* (2.5)*.*

(a) Input: $g \in L^2(\mathbb{S}^1 \times \mathbb{R})$.

(b) Compute $f_{\mathrm{FBP}} := (4\pi)^{-1}\mathbf{R}^* \circ \mathbf{I}^{-1} g$.

(c) Compute $\boldsymbol{\eta} := \mathbf{U}^* f_{\mathrm{FBP}}$ *via wavelet, curvelet or shearlet transform.*

(d) Apply rescaling $\boldsymbol{\eta} \leftarrow (2^{-j/2}\eta_\lambda)_{\lambda\in\Lambda}$.

(e) Output: Coefficients $\boldsymbol{\eta}$.

2.3 Inversion of the wave equation

We consider a planar geometry, which has been considered in our previous work [12]. Let $C_0^\infty(\mathbb{H}_+)$ denote the space of compactly supported functions $h\colon \mathbb{R}^2 \to \mathbb{R}$ that are supported in the half space $\mathbb{H}_+ := \mathbb{R} \times (0,\infty)$. For $f \in C_0^\infty(\mathbb{H}_+)$ consider the initial value problem

$$\begin{aligned} (\partial_t^2 - \Delta u(x,y,t) &= 0, && (x,y,t) \in \mathbb{R}^2 \times \mathbb{R} \\ u(x,y,0) &= y^{1/2} f(x,y) && (x,y) \in \mathbb{R}^2 \\ \partial_t u(x,y,0) &= 0 && (x,y) \in \mathbb{R}^2 \,. \end{aligned} \tag{2.8}$$

The trace map $\mathbf{A}\colon f \mapsto t^{-1/2} g$ where $g(x,t) := u(x,y=0,t)\chi\{t \geq 0\}$ for $(x,t) \in \mathbb{R}^2$ is known to be an isometry from $L^2(\mathbb{H}_+)$ to $L^2(\mathbb{H}_+)$, see [1, 12, 18]. In particular, the operator $\mathbf{A}$ is continuous.

Definition 2.8 (Forward operator for the wave equation)**.** *We define* $\mathbf{A}\colon L^2(\mathbb{H}_+) \to L^2(\mathbb{H}_+)$ *by* $\mathbf{A}f := t^{-1/2}u$*, for* $f \in C_0^\infty(\mathbb{H}_+)$*, where* u *is the solution of* (2.8)*, and extending it by continuity to* $L^2(\mathbb{H}_+)$*.*

The isometry property implies that any frame gives a DFD $(\mathbf{U}, \mathbf{V}, \boldsymbol{\kappa})$ by setting $v_\lambda = \mathbf{A}u_\lambda$ and $\kappa_\lambda = 1$. This, in particular, yields a wavelet vaguelette decomposition and a biorthogonal curvelet decomposition for the wave equation.

Example 2.9. *DFD for the wave equation Let* $(u_\lambda)_{\lambda\in\Lambda}$ *be either a wavelet frame, a curvelet frame or a shearlet frame with* $\lambda = (j,k,\beta) \in \Lambda$ *where* $j \geq 0$ *is the scale index. Then* $(\mathbf{U}, \mathbf{V}, \boldsymbol{\kappa})$ *is a DFD with*

$$v_\lambda := \mathbf{A}u_\lambda \tag{2.9}$$

$$\kappa_\lambda := 1\,. \tag{2.10}$$

As noted in [12] this result directly follows from the isometry property and the associated inversion formula $f = \mathbf{A}^*\mathbf{A}f$*.*

The isometry property also gives an efficient numerical algorithm for computing analysis coefficients with respect to the frame $\mathbf{V}$ in the case that $\mathbf{U}$ is associated with an efficient algorithm.

Algorithm 2.10. *Computing DFD coefficients for the wave equation Let* $\mathbf{U}$ *be the curvelet frame and define* $\mathbf{V}$ *by* (2.9).

(a) Input: $g \in L^2(\mathbb{H}_+)$.

(b) Compute $f_{\mathrm{FBP}} := \mathbf{A}^* g$.

(c) Compute $\boldsymbol{\eta} := \mathbf{U}^* f_{\mathrm{FBP}}$ *via wavelet, curvelet or shearlet transform.*

(d) Output: Coefficients $\boldsymbol{\eta}$.

The algorithm described above can be used for any problem where the forward operator $\mathbf{A}$ is an isometry. In the case of the wavelet transform this simple procedure has been previously used in [12].

3 Sparse ℓ^1-regularization

There are two fundamentally different and well-studied instances of sparse frame based regularization, namely ℓ^1-analysis regularization and ℓ^1-synthesis regularization. They are defined by

$$\mathbf{B}_\alpha^{\mathrm{ANA}}(\boldsymbol{y}_\delta) := \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{X}} \left\{ \frac{1}{2} \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}_\delta\|^2 + \alpha \sum_\lambda d_\lambda \, |\langle u_\lambda, \boldsymbol{x} \rangle| \right\} \tag{3.1}$$

$$\mathbf{B}_\alpha^{\mathrm{SYN}}(\boldsymbol{y}_\delta) := \mathbf{W} \left(\operatorname*{arg\,min}_{\boldsymbol{\xi} \in \ell^2(\Lambda)} \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{W}\boldsymbol{\xi} - \boldsymbol{y}_\delta\|^2 + \alpha \sum_\lambda d_\lambda \, |\xi_\lambda| \right\} \right) , \tag{3.2}$$

respectively, with weights $d_\lambda > 0$.

Definition 3.1. *We call* $\boldsymbol{\xi} \in \ell^2(\Lambda)$ *sparse if the set* $\{\lambda \in \Lambda \mid \xi_\lambda \neq 0\}$ *is finite.*

If $\boldsymbol{\xi} = (\xi_\lambda)_{\lambda \in \Lambda} \in \ell^2(\Lambda)$ is sparse, we write $\operatorname{Sign}(\boldsymbol{\xi}) := \{\boldsymbol{z} = (z_\lambda)_{\lambda \in \Lambda} \in \ell^2(\Lambda) \mid z_\lambda \in \operatorname{Sign}(\xi_\lambda)\}$ where $\operatorname{Sign}(\,\cdot\,) \colon \mathbb{R} \to \mathbb{R}$ is the multi-valued signum function defined by $\operatorname{Sign}(0) = [-1, 1]$ and $\operatorname{Sign}(x) = \{x / |x|\}$ for $x \neq 0$. We will use the notation

$$\|\cdot\|_{d,1} : \ell^2(\Lambda) \to \mathbb{R} \cup \{\infty\} \colon \boldsymbol{\xi} \mapsto \begin{cases} \sum_\lambda d_\lambda \, |\xi_\lambda| & \text{if } (d_\lambda \xi_\lambda)_{\lambda \in \Lambda} \in \ell^1(\Lambda) \\ \infty & \text{otherwise .} \end{cases} \tag{3.3}$$

Any element in the set $\operatorname{arg\,min}\{\|\mathbf{U}^*(\boldsymbol{x})\|_{d,1} \mid \mathbf{A}\boldsymbol{x} = \boldsymbol{y}\}$ is called $\|\mathbf{U}^*(\,\cdot\,)\|_{d,1}$-minimizing solution of $\mathbf{A}(\boldsymbol{x}) = \boldsymbol{y}$. Note that $\|\mathbf{U}^*(\,\cdot\,)\|_{d,1}$-minimizing solutions exists whenever there is any solution $\boldsymbol{x}$ with $\|\mathbf{U}^*(\boldsymbol{x})\|_{d,1} < \infty$, as follows from [22, Theorem 3.25].

Below we recall well-posedness and convergence results for both variants. These results hold under the following quite week assumptions:

(A1) $\mathbf{A}\colon \mathbb{X} \to \mathbb{Y}$ is bounded linear;

(A2) $\mathbf{U}, \mathbf{W}$ are synthesis operators of frames $(u_\lambda)_{\lambda\in\Lambda}$, $(w_\lambda)_{\lambda\in\Lambda}$ of $\mathbb{X}$;

(A3) $\boldsymbol{d} = (d_\lambda)_{\lambda\in\Lambda} \in \mathbb{R}^\Lambda$ satisfies $\inf\{d_\lambda \mid \lambda \in \Lambda\} > 0$.

For certain sparse elements we will state linear error estimates which have been derived in [13, 15]. See [6, 16, 20, 2, 22] for some further works on sparse ℓ^1-regularization, and [7, 17, 21] for wavelet regularization methods.

3.1 ℓ^1-Analysis regularization

Let us define the ℓ^1-analysis Tikhonov functional by

$$\mathcal{A}_{\alpha,\boldsymbol{y}}\colon \mathbb{X} \to \mathbb{R}\cup\{\infty\}\colon \boldsymbol{x} \mapsto \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \alpha\|\mathbf{U}^*\boldsymbol{x}\|_{\boldsymbol{d},1}\,. \tag{3.4}$$

Then we have $\mathbf{B}_\alpha^{\mathrm{ANA}}(\boldsymbol{y}) = \arg\min \mathcal{A}_{\alpha,\boldsymbol{y}}$.

Proposition 3.2 (Convergence of analysis regularization). *Let (A1)–(A3) be satisfied, suppose $\boldsymbol{y} \in \mathbb{Y}$, $\alpha > 0$, $(\boldsymbol{y}^k)_{k\in\mathbb{N}} \in \mathbb{Y}^\mathbb{N}$ with $\boldsymbol{y}^k \to \boldsymbol{y}$, and choose $\boldsymbol{x}^k \in \arg\min \mathcal{A}_{\alpha,\boldsymbol{y}^k}$.*

- EXISTENCE: *The functional $\mathcal{A}_{\alpha,\boldsymbol{y}}$ has at least one minimizer.*
- STABILITY: *There exists a subsequence $(\boldsymbol{x}^{k(\ell)})_{\ell\in\mathbb{N}}$ of $(\boldsymbol{x}^k)_{k\in\mathbb{N}}$ and a minimizer $\boldsymbol{x}_\alpha \in \arg\min \mathcal{A}_{\alpha,\boldsymbol{y}}$ such that $\|\boldsymbol{x}^{k(\ell)} - \boldsymbol{x}_\alpha\| \to 0$. If the minimizer $\boldsymbol{x}_\alpha$ of $\mathcal{A}_{\alpha,\boldsymbol{y}}$ is unique, then $\|\boldsymbol{x}^k - \boldsymbol{x}_\alpha\| \to 0$.*
- CONVERGENCE: *Assume $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{X}$ with $\|\mathbf{U}^*\boldsymbol{x}\|_{\boldsymbol{d},1} < \infty$ and suppose $\|\boldsymbol{y}^k - \boldsymbol{y}\| \leq \delta_k$ with $(\delta_k)_{k\in\mathbb{N}} \to 0$. Consider a parameter choice $(\alpha_k)_k \in (0,\infty)^\mathbb{N}$ such that $\lim_{k\to\infty}\alpha_k = \lim_{k\to\infty}\delta_k^2/\alpha_k = 0$. Then there is an $\|\mathbf{U}^*(\,\cdot\,)\|_{\boldsymbol{d},1}$-minimizing solution $\boldsymbol{x}^+$ of $\mathbf{A}(\boldsymbol{x}) = \boldsymbol{y}$ and a subsequence $(\boldsymbol{x}^{k(\ell)})_{\ell\in\mathbb{N}}$ with $\|\boldsymbol{x}^{k(\ell)} - \boldsymbol{x}^+\| \to 0$. If the $\|\mathbf{U}^*(\,\cdot\,)\|_{\boldsymbol{d},1}$-minimizing solution is unique, then $\|\boldsymbol{x}^k \to \boldsymbol{x}^+\| \to 0$.*

Proof. See [13, Propositions 5, 6 and 7]. □

In order to derive convergence rates, one has to make additional assumptions on the exact solution $\boldsymbol{x}^+$ to be recovered. Besides the sparsity this requires a certain interplay between $\boldsymbol{x}^+$ and the forward operator $\mathbf{A}$.

(A4) $\mathbf{U}^*\boldsymbol{x}^+$ is sparse;

(A5) $\exists \boldsymbol{z} \in \operatorname{Sign}(\mathbf{U}^*\boldsymbol{x}^+)\colon \mathbf{U}\boldsymbol{z} \in \operatorname{ran}(\mathbf{A}^*)$;

(A6) $\exists t \in (0,1)\colon \mathbf{A}$ is injective on $\operatorname{span}\{u_\lambda\colon |z_\lambda| > t\}$.

Assumption (A5) is a so-called source condition and the main restrictive assumption. It requires that there exists an element $z \in \operatorname{Sign}(\mathbf{U}^* x)$ that satisfies the smoothness assumption $\mathbf{U}z \in \operatorname{ran}(\mathbf{A}^*)$. Because $z \in \ell^2(\Lambda)$, the space $\operatorname{span}\{u_\lambda \colon |z_\lambda| > t\}$ is finite dimensional. Therefore, Condition (A6) requires injectivity on a certain finite dimensional subspace.

Proposition 3.3 (Convergence rates for analysis regularization). *Suppose (A1)–(A6) hold. Then, for a parameter choice $\alpha = C\delta$, there is a constant $c_+ \in (0, \infty)$ such that for all $y_\delta \in \mathbb{Y}$ with $\|\mathbf{A}x - y_\delta\| \leq \delta$ and every minimizer $x^{\alpha,\delta} \in \arg\min \mathcal{A}_{\alpha,y_\delta}$ we have $\|x^{\alpha,\delta} - x^+\| \leq c_+\delta$.*

Proof. See [15, Theorem III.8]. □

3.2 Synthesis regularization

Let use denote the ℓ^1-synthesis Tikhonov functional by

$$\mathcal{S}_{\alpha,y} \colon \ell^2(\Lambda) \to \mathbb{R} \cup \{\infty\} \colon x \mapsto \frac{1}{2}\|\mathbf{AW}\xi - y\|^2 + \alpha\|\xi\|_{d,1}. \tag{3.5}$$

Then it holds $\mathbf{B}_\alpha^{\mathrm{SYN}}(y) = \mathbf{W}(\arg\min \mathcal{S}_{\alpha,y})$. Synthesis regularization can be seen as analysis regularization for the coefficient inverse problem $\mathbf{AW}\xi = y$ and the analysis operator $\mathbf{U}^* = \mathrm{Id}$. Using Proposition 3.2 we therefore have the following result.

Proposition 3.4 (Convergence of synthesis regularization). *Let (A1)-(A3) be satisfied, suppose $y \in \mathbb{Y}$, $\alpha > 0$, $(y^k)_{k\in\mathbb{N}} \in \mathbb{Y}^{\mathbb{N}}$ with $y^k \to y$ and take $\xi^k \in \arg\min \mathcal{S}_{\alpha,y^k}$.*

- EXISTENCE: *The functional $\mathcal{S}_{\alpha,y}$ has at least one minimizer.*
- STABILITY: *There exists a subsequence $(\xi^{k(\ell)})_{\ell\in\mathbb{N}}$ of $(\xi^k)_{k\in\mathbb{N}}$ and $\xi_\alpha \in \arg\min \mathcal{S}_{\alpha,y}$ such that $(\xi^{k(\ell)})_\ell \to \xi_\alpha$. If the minimizer of $\mathcal{S}_{\alpha,y}$ is unique, then $\|\xi^k - \xi_\alpha\| \to 0$.*
- CONVERGENCE: *Assume $y = \mathbf{AW}\xi$ for $\xi \in \ell^2(\Lambda)$ with $\|\xi\|_{d,1} < \infty$ and $\|y^k - y\| \leq \delta_k$ with $(\delta_k)_{k\in\mathbb{N}} \to 0$. Consider a parameter choice $(\alpha_k)_{k\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}}$ with $\lim_{k\to\infty} \alpha_k = \lim_{k\to\infty} \delta_k^2/\alpha_k = 0$. Then there exist an $\|\cdot\|_{d,1}$-minimizing solution ξ^+ of $(\mathbf{AW})(\xi) = y$ and a subsequence $(\xi^{k(\ell)})_{\ell\in\mathbb{N}}$ with $\|\xi^{k(\ell)} - \xi^+\| \to 0$. If the $\|\cdot\|_{d,1}$-minimizing solution is unique, then $\|\xi^k - \xi^+\| \to 0$.*

Proof. Follows from Proposition 3.2 with $\mathbf{U} = \mathrm{Id}$ and $\mathbf{AW}$ in place of $\mathbf{A}$. □

We have linear convergence rates under the following additional assumptions on the element to be recovered.

(S4) $x^+ = \mathbf{W}\xi^+$ where $\xi^+ \in \ell^2(\Lambda)$ is sparse;

(S5) $\exists z \in \operatorname{Sign}(\xi^+) \colon z = \operatorname{ran}(\mathbf{W}^*\mathbf{A}^*)$;

(S6) $\exists t \in (0,1) \colon \mathbf{AW}$ is injective on $\operatorname{span}\{e_\lambda \colon |z_\alpha| > t\}$.

Proposition 3.5 (Convergence rates for synthesis regularization). *Suppose that (A1)–(A3) and (S4)–(S6) hold. Then, for a parameter choice $\alpha = C\delta$, there is a constant $c_+ \in (0, \infty)$ such that for all $y_\delta \in \mathbb{Y}$ with $\|\mathbf{A}x - y_\delta\| \leq \delta$, every minimizer $\xi^{\alpha,\delta} \in \arg\min \mathcal{S}_{\alpha,y_\delta}$ we have $\|\xi^{\alpha,\delta} - \xi^+\| \leq c_+\delta$.*

Proof. Follows from Proposition 3.3. □

Because $\mathbf{W}$ is bounded, the above convergence results can be transferred to convergence in the signal space $\mathbb{X}$. In particular, we have stability $\|\mathbf{W}(\xi^{k(\ell)}) - \mathbf{W}(\xi_\alpha)\| \to 0$ and convergence $\|\mathbf{W}\xi^{k(\ell)} - \mathbf{W}\xi^+\| \to 0$ under the assumptions made in Proposition 3.4, and linear convergence rates $\|\mathbf{W}\xi^{\alpha,\delta} - x^+\| \leq \tilde{c}_+\delta$ under the assumptions made in Proposition 3.5.

3.3 Sparse regularization using an SVD

In the special case that $\mathbf{U}$ is part of an SVD, then analysis and synthesis regularization are equivalent and can be computed explicitly by soft-thresholding of the expansion coefficients.

Theorem 3.6 (Equivalence in the SVD case). *Let $(\mathbf{U}, \mathbf{V}, \kappa)$ be an SVD for $\mathbf{A}$, let $y_\delta \in \mathbb{Y}$ and consider (3.1), (3.2) with $\mathbf{W} = \mathbf{U}$. Then*

$$\mathbf{B}_\alpha^{\mathrm{ANA}}(y_\delta) = \mathbf{B}_\alpha^{\mathrm{SYN}}(y_\delta) = \{(\mathbf{U} \circ \mathbf{M}_\kappa^+ \circ \mathbb{S}_{\alpha d/\kappa} \circ \mathbf{V}^*)(y_\delta)\}\,,$$

equals the soft-thresholding estimator in the SVD system.

Proof. Because $\mathbf{U}$ is an orthonormal basis of $\mathbb{X}$, we have $x = \mathbf{U}\xi \Leftrightarrow \xi = \mathbf{U}^* x$ which implies that $\mathbf{B}_\alpha^{\mathrm{ANA}}(y_\delta) = \mathbf{B}_\alpha^{\mathrm{SYN}}(y_\delta)$. Now let $x_\alpha \in \mathbf{B}_\alpha^{\mathrm{SYN}}(y_\delta)$ be any minimizer of the ℓ^1-analysis Tikhonov functional $\mathcal{A}_{\alpha,y}$. Let $\mathbf{P}_{\mathrm{ran}(\mathbf{A})^\perp}$ denote the orthogonal projection on $\mathrm{ran}(\mathbf{A})^\perp$. We have

$$\begin{aligned}
\mathcal{A}_{\alpha,y}(x) &= \frac{1}{2}\|\mathbf{A}x - y\|^2 + \alpha\|\mathbf{U}^* x\|_{d,1} \\
&= \left\|\mathbf{P}_{\mathrm{ran}(\mathbf{A})^\perp}(y)\right\|^2 + \sum_{\lambda\in\Lambda} \frac{1}{2}\left|\langle \mathbf{A}x - y, v_\lambda\rangle\right|^2 + \sum_{\lambda\in\Lambda} \alpha d_\lambda \left|\langle x, u_\lambda\rangle\right| \\
&= \left\|\mathbf{P}_{\mathrm{ran}(\mathbf{A})^\perp}(y)\right\|^2 + \sum_{\lambda\in\Lambda} \frac{1}{2}\left|\kappa_\lambda \langle x, u_\lambda\rangle - \langle y, v_\lambda\rangle\right|^2 + \alpha d_\lambda \left|\langle x, u_\lambda\rangle\right| .
\end{aligned}$$

The latter sum is minimized by componentwise soft-thresholding. This shows $x_{\alpha,\delta} = (\mathbf{U} \circ \mathbb{S}_{\alpha d/\kappa^2} \circ \mathbf{M}_\kappa^+ \circ \mathbf{V}^*)(y_\delta) = (\mathbf{U} \circ \mathbf{M}_\kappa^+ \circ \mathbb{S}_{\alpha d/\kappa} \circ \mathbf{V}^*)(y_\delta)$ and concludes the proof. □

In the case that $(\mathbf{U}, \mathbf{V}, \kappa)$ is a redundant DFD expansion and not an SVD, then (3.1), (3.2), and the soft-thresholding estimator

$$(\mathbf{U} \circ \mathbf{M}_\kappa^+ \circ \mathbb{S}_{\alpha d/\kappa} \circ \mathbf{V}^*)(y_\delta) \tag{3.6}$$

are all non-equivalent. Further, in this case, (3.1) and (3.2) have to be computed by iterative minimization algorithms. This requires repeated application of the forward and adjoint problem and therefore is time consuming. In the following section, we study

DFD thresholding which is the analog of (3.6) for redundant systems. Despite the non-equivalence to ℓ^1-regularization, we are able to derive the same type of convergence results and linear convergence rates as for the analysis and synthesis variants of ℓ^1-regularization.

4 Regularization via DFD thresholding

Throughout this section we fix the following assumptions:

(B1) $\mathbf{A}\colon \mathbb{X} \to \mathbb{Y}$ is bounded linear.

(B2) $(\mathbf{U}, \mathbf{V}, \boldsymbol{\kappa})$ is s DFD for $\mathbf{A}$.

(B3) $\boldsymbol{d} = (d_\lambda)_{\lambda\in\Lambda} \in \mathbb{R}^\Lambda$ satisfies $\inf\{d_\lambda \mid \lambda \in \Lambda\} > 0$.

In this section we show well-posedness, convergence and convergence rates for DFD soft-thresholding.

4.1 DFD soft-thresholding

Any DFD gives an explicit inversion formula $\boldsymbol{x} = (\bar{\mathbf{U}} \circ \mathbf{M}_{\boldsymbol{\kappa}}^{+} \circ \mathbf{V}^*)(\mathbf{A}\boldsymbol{x})$ where $\mathbf{M}_{\boldsymbol{\kappa}}^{+}$ is defined by (2.2). For ill-posed problems, $\kappa_\lambda \to 0$ and therefore the above reproducing formula is unstable when applied to noisy data $\boldsymbol{y}_\delta$ instead of $\mathbf{A}\boldsymbol{x}$. Below we stabilize the inversion by including the soft-thresholding operation.

Definition 4.1 (DFD soft-thresholding). *Let $(\mathbf{U}, \mathbf{V}, \boldsymbol{\kappa})$ be a DFD for $\mathbf{A}$. We define the nonlinear DFD soft-thresholding estimator by*

$$\mathbf{B}_\alpha^{\mathrm{DFD}}\colon \mathbb{Y} \to \mathbb{X}\colon \boldsymbol{y} \mapsto (\bar{\mathbf{U}} \circ \mathbf{M}_{\boldsymbol{\kappa}}^{+} \circ \mathbb{S}_{\alpha\boldsymbol{d}/\boldsymbol{\kappa}} \circ \mathbf{V}^*)(\boldsymbol{y}). \tag{4.1}$$

If $(u_\lambda)_\lambda$, $(v_\lambda)_\lambda$ are ONBs, then Theorem 3.6 shows that (3.1) and (3.2) are equivalent to (4.1). In the case of general frames, $\mathbf{B}_\alpha^{\mathrm{ANA}}$, $\mathbf{B}_\alpha^{\mathrm{SYN}}$ and $\mathbf{B}_\alpha^{\mathrm{DFD}}$ are all different.

As the main result in this paper we show that DFD soft-thresholding yields the same theoretical results as ℓ^1-regularization. Assuming efficient implementations for $\bar{\mathbf{U}}$ and $\mathbf{V}^*$, the DFD estimator has the advantage that it can be calculated non-iteratively and is therefore much faster than $\mathbf{B}_\alpha^{\mathrm{SYN}}$ and $\mathbf{B}_\alpha^{\mathrm{DFD}}$.

Consider the ℓ^1-Tikhonov functional for the multiplication operator $\mathbf{M}_{\boldsymbol{\kappa}}$,

$$\mathcal{M}_{\alpha,\boldsymbol{\eta}}\colon \ell^2(\Lambda) \to \mathbb{R} \cup \{\infty\}\colon \boldsymbol{\xi} \mapsto \frac{1}{2}\|\mathbf{M}_{\boldsymbol{\kappa}}\boldsymbol{\xi} - \boldsymbol{\eta}\|^2 + \alpha\|\boldsymbol{\xi}\|_{\boldsymbol{d},1}. \tag{4.2}$$

The proof strategy used in this paper is based on the following Lemma.

Lemma 4.2 (ℓ^1-minimization for multiplication operators).

(a) $\forall \alpha \in \mathbb{R}_{>0} \forall \boldsymbol{\eta} \in \ell^2(\Lambda)$: *$\mathcal{M}_{\alpha,\boldsymbol{\eta}}$ has a unique minimizer.*

(b) $\forall \alpha \in \mathbb{R}_{>0} \forall \boldsymbol{\eta} \in \ell^2(\Lambda)$: $(\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}})(\boldsymbol{\eta}) = \arg\min \mathcal{M}_{\alpha,\boldsymbol{\eta}}$.

(c) $\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}} \colon \ell^2(\Lambda) \to \ell^2(\Lambda)$ *is continuous.*

(d) $\mathbf{V}^*\mathbf{A} = \mathbf{M}_{\boldsymbol{\kappa}}\mathbf{U}^*$.

Proof. Because $(\mathrm{Id}, \mathrm{Id}, \boldsymbol{\kappa})$ is an SVD for $\mathbf{M}_{\boldsymbol{\kappa}}$, Items (a), (b) follow from Theorem 3.6, the equivalence of ℓ^1-regularization and soft-thresholding in the SVD case. Item (c) follows from Proposition 3.4. Moreover, the equality $(\mathbf{V}^*\mathbf{A}\boldsymbol{x})_\lambda = \langle v_\lambda, \mathbf{A}\boldsymbol{x}\rangle = \langle \mathbf{A}^* v_\lambda, \boldsymbol{x}\rangle = \kappa_\lambda \langle u_\lambda, \boldsymbol{x}\rangle = (\mathbf{M}_{\boldsymbol{\kappa}}\mathbf{U}^*\boldsymbol{x})_\lambda$ shows Item (d). □

Note that the continuity of $\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}}$ (see Item (b) in the above lemma) is not obvious as it is the composition of the soft thresholding $\mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}}$ with the discontinuous operator in $\mathbf{M}_{\boldsymbol{\kappa}}^+$. The characterization in Item (b) of $\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}}$ as minimizer of the ℓ^1-Tikhonov functional $\mathcal{M}_{\alpha,\eta}$ and the existing stability results for ℓ^1-Tikhonov regularization yields an elegant way to obtain the continuity of $\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}}$. Verifying the continuity directly would also be possible but seems to be a harder task. A similar comment applies to the proof of Theorem 4.3 where we use the convergence of $\mathcal{M}_{\alpha,\eta}$ to show convergence the DFD soft-thresholding estimator $\mathbf{B}_\alpha^{\mathrm{DFD}}$.

4.2 Convergence analysis

In this section, we show that $(\mathbf{B}_\alpha^{\mathrm{DFD}})_{\alpha>0}$ is well-posed and convergent.

Theorem 4.3 (Well-posedness and convergence). *Let (B1)-(B3) be satisfied, suppose $\boldsymbol{y} \in \mathbb{Y}$ and let $(\boldsymbol{y}^k)_{k\in\mathbb{N}} \in \mathbb{Y}^{\mathbb{N}}$ satisfy $\boldsymbol{y}^k \to \boldsymbol{y}$.*

(a) Existence: $\mathbf{B}_\alpha^{\mathrm{DFD}} \colon \mathbb{Y} \to \mathbb{X}$ *is well-defined for all* $\alpha > 0$.

(b) Stability: $\mathbf{B}_\alpha^{\mathrm{DFD}} \colon \mathbb{Y} \to \mathbb{X}$ *is continuous for all* $\alpha > 0$.

(c) Convergence: *Assume $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$ for some $\boldsymbol{x} \in \mathbb{X}$ with $\|\mathbf{U}^*\boldsymbol{x}\|_{\boldsymbol{d},1} < \infty$, suppose $\|\boldsymbol{y}^k - \boldsymbol{y}\| \le \delta_k$ with $(\delta_k)_{k\in\mathbb{N}} \to 0$ and consider a parameter choice $(\alpha_k)_{k\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}}$ with $\lim_{k\to\infty} \alpha_k = \lim_{k\to\infty} \delta_k^2/\alpha_k = 0$. Then $\|\mathbf{B}_{\alpha_k}^{\mathrm{DFD}}(\boldsymbol{x}^k) - \boldsymbol{x}^+\| \to 0$.*

Proof. (a), (b): According to Lemma 4.2, the mapping $\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}} \colon \ell^2(\Lambda) \to \ell^2(\Lambda)$ is well-defined and continuous. Moreover, by definition we have $\mathbf{B}_\alpha^{\mathrm{DFD}} = \bar{\mathbf{U}} \circ (\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}}) \circ \mathbf{V}^*$ which implies existence and stability of DFD thresholding.
(c): We have

$$\begin{aligned}\|\mathbf{B}_{\alpha_k}^{\mathrm{DFD}}(\boldsymbol{y}^k) - \boldsymbol{x}^+\| &= \|\bar{\mathbf{U}} \circ (\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}}) \circ \mathbf{V}^*(\boldsymbol{y}^k) - \bar{\mathbf{U}}\mathbf{U}^*\boldsymbol{x}^+\| \\ &\le \|\mathbf{U}\| \, \|(\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}})(\mathbf{V}^*\boldsymbol{y}^k) - \mathbf{U}^*\boldsymbol{x}^+\|\,. \end{aligned} \tag{4.3}$$

Moreover, $\|\mathbf{V}^*\boldsymbol{y}^k - \mathbf{M}_{\boldsymbol{\kappa}}\mathbf{U}^*\boldsymbol{x}^+\| = \|\mathbf{V}^*\boldsymbol{y}^k - \mathbf{V}^*\mathbf{A}\boldsymbol{x}^+\| \le \|\mathbf{V}\|\|\boldsymbol{y}^k - \mathbf{A}\boldsymbol{x}^+\| \le \|\mathbf{V}\|\delta_k$. Therefore, Proposition 3.4 and the equality $\arg\min \mathcal{M}_{\alpha,\mathbf{V}^*\boldsymbol{y}^k} = (\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}})(\mathbf{V}^*\boldsymbol{y}^k)$ shown in Lemma 4.2 imply $\|(\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha \boldsymbol{d}/\boldsymbol{\kappa}})(\mathbf{V}^*\boldsymbol{y}^k) - \mathbf{U}^*\boldsymbol{x}^+\| \to 0$ for $k \to \infty$. Together with (4.3) this yields (c) and completes the proof. □

4.3 Convergence rates

Next we derive linear convergence rates for sparse solutions. Let use denote by $\operatorname{supp}(\xi) := \{\lambda \in \Lambda \colon \xi_\lambda \neq 0\}$ the support of $\xi \in \ell^2(\Lambda)$. To derive the convergence rates, we assume the following for the exact solution $\boldsymbol{x}^+$ to be recovered.

(B4) $\mathbf{U}^*\boldsymbol{x}^+$ is sparse.

(B5) $\forall \lambda \in \operatorname{supp}(\mathbf{U}^*\boldsymbol{x}^+) \colon \kappa_\lambda \neq 0$.

Note that assumptions (B4), (B5) imply the source condition

$$\boldsymbol{z} \in \partial \| \cdot \|_{\boldsymbol{d},1} \cap \operatorname{ran}(\mathbf{M}_{\boldsymbol{\kappa}}^*) \neq \emptyset .$$

is satisfied for some element $\boldsymbol{z} \in \ell^2(\Lambda)$ that can be chosen such that $|z_\lambda| < 1$ for $\lambda \notin \operatorname{supp}(\mathbf{U}^*\boldsymbol{x}^+)$. Moreover, it follows that $\mathbf{M}_{\boldsymbol{\kappa}}$ is injective on $\operatorname{span}\{e_\lambda \mid |z_\lambda| > t\}$ with $t := \max\{|\kappa_\lambda| \mid \lambda \notin \operatorname{supp}(\mathbf{U}^*\boldsymbol{x}^+)\}$. Because $\mathbf{U}^*\boldsymbol{x}^+ \in \ell^2(\Lambda)$, we have $t < 1$.

Theorem 4.4 (Convergence rates)**.** *Suppose that (B1)–(B5) hold. Then, for the parameter choice $\alpha = C\delta$ with $C \in (0,\infty)$, there is a constant $c_+ \in (0,\infty)$ such that for all $\boldsymbol{y}_\delta \in \mathbb{Y}$ with $\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}_\delta\| \leq \delta$ we have $\|\mathbf{B}_\alpha^{\mathrm{DFD}}(\boldsymbol{y}_\delta) - \boldsymbol{x}^+\| \leq c_+\delta$.*

Proof. As in the proof of Theorem 4.3 one obtains

$$\|\mathbf{B}_\alpha^{\mathrm{DFD}}(\boldsymbol{y}^k) - \boldsymbol{x}^+\| \leq \|\mathbf{U}\| \, \|(\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha\boldsymbol{d}/\boldsymbol{\kappa}})(\mathbf{V}^*\boldsymbol{y}^k) - \mathbf{U}^*\boldsymbol{x}^+\| \tag{4.4}$$

$$\|\mathbf{V}^*\boldsymbol{y}^k - \mathbf{M}_\lambda \mathbf{U}^*\boldsymbol{x}^+\| \leq \|\mathbf{V}\|\delta . \tag{4.5}$$

According to the considerations below (B4), (B5) the conditions (S4)–(S6) are satisfied for the operator $\mathbf{M}_{\boldsymbol{\kappa}}$ in place of $\mathbf{A}$ and with $\mathbf{W} = \mathrm{Id}$. The convergence rates result in Proposition 3.3, estimate (4.5), and the identity $\arg\min \mathcal{M}_{\alpha,\mathbf{V}^*\boldsymbol{y}} = (\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha\boldsymbol{d}/\boldsymbol{\kappa}})(\mathbf{V}^*\boldsymbol{y})$ shown in Lemma 4.2 imply $\|(\mathbf{M}_{\boldsymbol{\kappa}}^+ \circ \mathbb{S}_{\alpha\boldsymbol{d}/\boldsymbol{\kappa}})(\mathbf{V}^*\boldsymbol{y}) - \mathbf{U}^*\boldsymbol{x}^+\| \leq c\delta$. Together with (4.4) this implies $\|\mathbf{B}_\alpha^{\mathrm{DFD}}(\boldsymbol{y}_\delta) - \boldsymbol{x}^+\| \leq c\|\mathbf{U}\|\delta$ and concludes the proof. □

5 Conclusion

To overcome the inherent ill-posedness of inverse problems, regularization methods incorporate available prior information about the unknowns to be reconstructed. In this context, a useful prior is sparsity with respect to a certain frame. There are at least two different regularization strategies implementing sparsity with respect to a frame, namely ℓ^1-analysis regularization and ℓ^1-synthesis regularization. In this paper, we analyzed DFD-thresholding as a third variant of sparse regularization. One advantage of DFD-thresholding compared to other sparse regularization methods is its non-iterative nature leading to fast algorithms. Besides having a DFD, actually computing the DFD soft-thresholding estimator (4.1) requires the dual frame $\bar{\mathbf{U}}$. While in the general situation, the DFD and the dual frame have to be computed numerically, we have shown that for many practical examples (see Section 2) they are known explicitly and efficient algorithms are available for its numerical evaluation.

The DFD-approach presented in this paper is well studied in the context of statistical estimating using certain multi-scale systems. However, its analysis in the context of regularization theory has not been given so far. In this paper we closed this gap and presented a complete convergence analysis of DFD-thresholding as regularization method.

Acknowledgement

The work of M.H has been supported by the Austrian Science Fund (FWF), project P 30747-N32.

References

[1] A. L. Buhgeim and V. B. Kardakov. Solution of an inverse problem for an elastic wave equation by the method of spherical means. *Sibirsk. Mat. Z.*, 19(4):749–758, 1978.

[2] S. Bürger, J. Flemming, and B. Hofmann. On complex-valued deautoconvolution of compactly supported functions with sparse fourier representation. *Inverse Probl.*, 32(10):104006, 2016.

[3] E. J. Candès and D. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *Ann. Statist.*, 30(3):784–842, 2002.

[4] O. Christensen. *An introduction to frames and Riesz bases.* Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2003.

[5] F. Colonna, G. Easley, K. Guo, and D. Labate. Radon transform inversion using the shearlet representation. *Appl. Comput. Harmon. Anal.*, 29(2):232–250, 2010.

[6] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[7] V. Dicken and P. Maass. Wavelet-Galerkin methods for ill-posed problems. *J. Inverse Ill-Posed Probl.*, 4:203–221, 1996.

[8] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126, 1995.

[9] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inveerse Problems*, 23(3):947, 2007.

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

[11] J. Frikel. Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, 34(1):117–141, 2013.

[12] J. Frikel and M. Haltmeier. Efficient regularization with wavelet sparsity constraints in photoacoustic tomography. *Inverse Problems*, 34(2):024006, 2018.

[13] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Problems*, 24(5):055020, 13, 2008.

[14] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.

[15] M. Haltmeier. Stable signal reconstruction via ℓ^1-Minimization in redundant, non-tight frames. *IEEE Trans. Sig. Processing*, 61(2):420–426, 2013.

[16] D. A. Lorenz. Convergence rates and source conditions for tikhonov regularization with sparsity constraints. *Journal of Inverse and Ill-Posed Problems*, 16(5):463–478, 2008.

[17] A.K. Louis, P. Maass, and A. Rieder. *Wavelets. Theorie und Anwendungen.* Teubner, Stuttgart, 1998.

[18] E. K. Narayanan and Rakesh. Spherical means with centers on a hyperplane in even dimensions. *Inverse Probl.*, 26(3):035014, March 2010.

[19] F. Natterer. *The Mathematics of Computerized Tomography*, volume 32 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 2001.

[20] R. Ramlau and G. Teschke. A tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numerische Mathematik*, 104(2):177–203, 2006.

[21] A. Rieder. A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization. *Numer. Math.*, 75:501–522, 1997.

[22] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. Variational methods in imaging. *Applied Mathematical Sciences*, 167, 2009.