Nr.01 / 2013

Leopold-Franzens-Universität Innsbruck

Preprint-Series: Department of Mathematics - Applied Mathematics

Stable Signal Reconstruction via ℓ^1-Minimization in Redundant, Non Tight-Frames

Markus Haltmeier

APPLIEDMATHEMATICS

Technikerstraße 21a - 6020 Innsbruck - Austria
Tel.: +43 512 507 53803 Fax: +43 512 507 53898
https:\\www.applied-math.uibk.ac.at

Stable Signal Reconstruction via ℓ^1-Minimization in Redundant, Non-Tight Frames

Markus Haltmeier

***Abstract*—In many signal and image processing applications a desired clean signal is distorted from blur and noise. Reconstructing the clean signal usually yields to a high dimensional ill-conditioned system of equations, where a direct solution would severely amplify the noise. Stable signal reconstruction requires the use of regularization techniques, which incorporate a-priori knowledge about the signal. A particular successful property for that purpose is the sparsity of the analysis coefficients of the clean signal in a suitable frame or dictionary, which can be implemented via ℓ^1-minimization. Most existing stable recovery results for ℓ^1-analysis minimization require the frame to be an orthonormal basis. This contrasts practical applications, where redundant frames often perform better than bases. In this paper we address this issue and derive stable recovery results for ℓ^1-analysis minimization in redundant, possibly non-tight frames.**

***Index Terms*—Sparsity, inverse problems, ℓ^1-minimization, stable signal reconstruction, analysis prior, sparse recovery, frames, compressed sensing.**

I. Introduction

NUMEROUS applications in signal and image processing require the approximate solution of a linear ill-conditioned or ill-posed equation

$$\mathbf{K}u = v_\delta \qquad \text{for given} \qquad v_\delta \in \mathcal{V}\,. \tag{I.1}$$

Here $\mathbf{K}\colon \mathcal{U} \to \mathcal{V}$ is a bounded linear mapping between Hilbert spaces $\mathcal{U}$ and $\mathcal{V}$ modeling the data acquisition process and v_δ are available noisy data. We assume bounded noise, which means that the given data satisfy the estimate $\|v_\delta - \mathbf{K}u_\star\| \leq \delta$ for some unknown noise-free data $\mathbf{K}u_\star$ and some known bound $\delta > 0$ for the noise level. The aim is to reconstruct an approximation to $u_\star$ based on the available data v_δ and the noise bound δ.

The equation (I.1) is called ill-posed, if the operator $\mathbf{K}$ is not continuously invertible. The prototype of such an ill-posed problem is deconvolution, where $(\mathbf{K}u)(x) = (u * k)(x)$ is the convolution of a signal $u \in L^2(\mathbb{R})$ with a smooth kernel function $k\colon \mathbb{R} \to \mathbb{R}$. In such a case, the exact solution of (I.1) may not exist (if $\operatorname{Ran}(\mathbf{K}) \neq \mathcal{V}$), may not be unique (if $\operatorname{Ker}(\mathbf{K}) \neq \{0\}$), or may be arbitrarily far away from the true solution $u_\star$ (if $\mathbf{K}^{-1}$ is discontinuous). Discretizing an ill-posed problem yields to an ill-conditioned system of equations, where direct solutions significantly amplify noise in the data. In oder to make the signal reconstruction process stable, regularization methods have to be applied, which use a-priori knowledge about the true unknown $u_\star$ in order to construct approximate but stable solutions of (I.1) that converge to $u_\star$ as $\delta \to 0$.

In this paper we assume that the analysis coefficients of the clean signal $u_\star$ with respect to a redundant frame have small ℓ^1-norm. We derive linear error estimates $\|u_\star - u_\delta\| \leq c\delta$, with some constant $c \in (0, \infty)$, between the true signal $u_\star$ and the reconstructed signal u_δ from noisy data obtained by ℓ^1-minimization. Our main results (stated in Theorem III.8) are formulated in a possible infinite dimensional setting. These results, in particular, apply to finite dimensional signal spaces, where the linear operator may be considered as matrix $\mathbf{K} \in \mathbb{R}^{d \times n}$, that defines an (essentially) underdetermined large system of linear equations.

A. Regularization Methods

Regularization methods have to account for both, the non-uniqueness and the instability of solving (I.1). Classical methods address these issues by forcing solutions to have small quadratic Hilbert space norm. A common approach is $\|\cdot\|^2$-constrained regularization

$$\begin{aligned} &\text{minimize} && \|u\|^2 \\ &\text{such that} && \|\mathbf{K}u - v_\delta\| \leq \delta\,. \end{aligned} \tag{I.2}$$

Another closely related approach is to define regularized solutions $u_{\alpha,\delta}$ as minimizers of the classical Tikhonov functional

$$\|\mathbf{K}u - v_\delta\|^2 + \alpha\|u\|^2\,. \tag{I.3}$$

where α is a properly chosen regularization parameter. Note that (I.3) and (I.2) are equivalent if the parameter α is selected according to Morozov's discrepancy principle, which chooses the largest $\alpha = \alpha(v_\delta, \delta)$ such that $\|\mathbf{K}u_{\alpha,\delta} - v_\delta\| = \delta$. However, Tikhonov regularization can also be applied with other parameter choice rules in which case the two reconstruction methods are not equivalent.

Standard results of regularization theory [1], [2] then show that (I.2) and (I.3) have unique minimizers which strongly converge to the minimal norm solution of the equation $\mathbf{K}u = v$ as the noise level tends to zero. However these classical Hilbert space methods suffer from two shortcomings.

1) In the case that the operator $\mathbf{K}$ is not injective, among all solutions of the noise free equation $\mathbf{K}u = v$, the one with smallest Hilbert space norm $\|\cdot\|$ is approximated. Several recent applications (probably the most prominent being compressed sensing) show that these solution are often not the relevant ones in practice.
2) In the case that the inverse of $\mathbf{K}$ is discontinuous (or has large norm), the convergence of the regularized solutions

M. Haltmeier is with the Department of Mathematics at the University of Innsbruck, Technikerstraße 21a/2, 6020 Innsbruck, Austria (ma.haltmeier@gmail.com).

u_δ to the minimal norm solution $u_\star$ is slow. In order to get reasonable error estimates the true signal has to satisfy additional assumptions, such as $u_\star = \mathbf{K}^T \lambda$ for some $\lambda \in \mathcal{V}$. Good error estimates require small value of $\|\lambda\|$. For severely ill-posed problems (such as deconvolution) this is a strong smoothness assumption on $u_\star$. As a result, only low-resolution signals are recovered accurately.

Recent studies demonstrate that both shortcomings of quadratic regularization can be avoided by replacing the quadratic penalty $\|\cdot\|^2$ with non-differentiable penalties, such as the ℓ^1-norm or the total variation, which can provide accurate approximations to high resolution signals.

B. Sparse ℓ^1-Regularization

In this paper we focus on regularization methods based on the ℓ^1-norm of the analysis coefficients with respect to some prescribed frame $\mathcal{D} = \{\phi_\omega : \omega \in \Omega\}$ of the Hilbert space $\mathcal{U}$. We denote by $\mathbf{\Phi}^T u = (\langle \phi_\omega, u \rangle : \omega \in \Omega)$ the analysis operator corresponding to $\mathcal{D}$ and aim at finding approximate solutions of (I.1) having small value of

$$\|\mathbf{\Phi}^T u\|_1 = \sum_{\omega \in \Omega} |\langle \phi_\omega, u \rangle| \,. \tag{I.4}$$

This accounts for the fact that signals of practical interest are often much better represented by solutions having small $\|\mathbf{\Phi}^T u\|_1$ instead of solutions with small Hilbert space norm.

The first method we will analyze is ℓ^1-constrained regularization, which replaces the quadratic Hilbert space norm $\|u\|^2$ in (I.2) by $\|\mathbf{\Phi}^T u\|_1$ and considers solutions of

$$\begin{aligned} &\text{minimize} \;\; \|\mathbf{\Phi}^T u\|_1 \\ &\text{such that} \;\; \|\mathbf{K}u - v_\delta\| \leq \delta \,. \end{aligned} \tag{I.5}$$

We will also study the following unconstrained version of (I.5), which we will refer to as ℓ^1-Tikhonov regularization. There, regularized solutions are defined as minimizers of

$$\mathrm{T}_{v_\delta, \alpha}(u) := \|\mathbf{K}u - v_\delta\|^2 + \alpha \|\mathbf{\Phi}^T u\|_1 \,. \tag{I.6}$$

Again, the minimization problems (I.6) and (I.5) are equivalent if the regularization parameter α in Tikhonov regularization is chosen according to Morozov's discrepancy principle, but differ for other parameter choice rules.

Assume for the moment that the family $\mathcal{D}$ is an orthonormal basis of the space $\mathcal{U}$. Then, under the assumptions that the source condition $\mathbf{\Phi} \circ \partial \|\cdot\|_1 (u_\star) \cap \mathrm{Ran}\left(\mathbf{K}^T\right) \neq \emptyset$ is satisfied (here $\partial\|\cdot\|_1$ denotes the subdifferential of the ℓ^1-norm; see Lemma II.3) and that $\mathbf{K}$ is injective on a certain finite dimensional subspace of $\mathcal{U}$, in [3], [4] we derived the linear error estimates $\|u_\star - u_\delta\| \leq c\delta$ for the minimizers u_δ of either (I.6) and (I.5). Hence, solving the ill-posed equation by ℓ^1-minimization provides the same linear error estimate as a well posed problem would offer. In the orthonormal basis case, linear error estimates for ℓ^1-minimization appeared earlier in the compressed sensing (or compressive sampling) context in [5]. These results are stronger in the sense that the estimates there hold *uniformly* for all sufficiently sparse solutions. However, the results of [5] require the stronger assumption that the operator $\mathbf{K}$ satisfies the so called restricted isometry property (RIP). Indeed, as shown in [3], even in a possibly infinite dimensional setting, the RIP implies the source condition uniformly for all sufficiently sparse elements. The RIP is known to be satisfied for many random matrices (with high probability) and is therefore of high importance for compressed compressed sensing, where the sensing matrix can be designed at last with some degree of freedom. In contrast, in typical inverse problems, such as in deconvolution, the operator $\mathbf{K}$ is non-random, often specified by the specific application and does not satisfy the restricted isometry property. In such a situation the estimates of [3], [4] still provide unique and stable recovery for certain sparse elements.

The error estimates of [3], [4] require the family $\mathcal{D}$ to be an orthonormal basis. However, in typical signal and image processing applications much better reconstruction results are usually obtained when redundant systems (such as translation invariant wavelets) are used instead of orthonormal ones (such as orthonormal wavelets). The goal of this paper is to close the gap between theory and application and to derive linear error estimates for ℓ^1-regularization in redundant, possibly non-tight frames.

C. Analysis vs Synthesis Prior

The minimization problems (I.5) and (I.6) are designed to provide stable approximations to solutions of the constrained minimization problem

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{\Phi}^T u\|_1 \\ &\text{such that} \quad \mathbf{K}u = v \,. \end{aligned} \tag{I.7}$$

The solutions of the minimization problem (I.7) are those solutions of the equation $\mathbf{K}u = v$ whose analysis coefficients $\mathbf{\Phi}^T u$ have minimal ℓ^1-norm. A different solution approach would be to take another frame $\{\psi_\omega : \omega \in \Omega\}$ with synthesis operator $\mathbf{\Psi}x = \sum_{\omega \in \Omega} \psi_\omega x(\omega)$ and to consider solutions of the equation $\mathbf{K}u = v$ that are synthesized in the new frame by vectors with small ℓ^1-norm. More precisely, one defines $u_\star = \mathbf{\Psi} x_\star$, where $x_\star$ is a solution of

$$\begin{aligned} &\text{minimize} \quad \|x\|_1 \\ &\text{such that} \quad \mathbf{K}\mathbf{\Psi}x = v \,. \end{aligned} \tag{I.8}$$

Again one can define stable versions of this minimization problem by either a constrained approach or a penalty approach. According to the terminology of [6], regularization methods based on (I.8) use an ℓ^1-synthesis prior, whereas regularization methods based on (I.7) use an ℓ^1-analysis prior.

In that case that $\mathcal{D}$ is a basis, then the choice $\mathbf{\Psi} = (\mathbf{\Phi}^T)^{-1}$ makes the analysis approach and the synthesis approach equivalent. In this case, one readily verifies that $u_\star$ is a minimizer of (I.7) if and only if $x_\star = \mathbf{\Phi}^T u_\star$ is a minimizer of (I.8). If $\mathcal{D}$ is a redundant dictionary, however, the two approaches are different, as discussed extensively in [6], [7].

It may depend on a particular situation which of the two approaches should be favored. For the solution of ill-posed problems, however, the analysis approach seems to be more natural for at least two reasons. First, the synthesis approach should solve the ill-posed equation $\mathbf{K}u = v$ and at the same time would solve a sparse approximation problem for the potential solution. This may be a too ambitious goal. Second, the parameter space $\ell^2(\Omega)$ is much higher dimensional then the signal space $\mathcal{U}$, and therefore the new equation $\mathbf{K}\mathbf{\Psi}x = v$ considered in the synthesis approach (I.8) is even more ill-posed and underdetermined than the original one.

D. Main Results

As the main results of this paper, we derive sufficient conditions that guarantee linear error estimates for ℓ^1-analysis regularization in redundant frames using the unconstrained program (I.5) or the unconstrained version (I.6). Roughly spoken our results are as follows. Under the assumptions that

1) $\mathbf{\Phi}^T u_\star$ is a sparse sequence,
2) $\mathbf{\Phi} \circ \partial\| \cdot \|_1(\mathbf{\Phi}^T u_\star) \cap \operatorname{Ran}\left(\mathbf{K}^T\right) \neq \emptyset$,
3) $\mathbf{K}$ is injective on a certain finite dimensional subspace $\mathcal{U}_\star \subset \mathcal{U}$,

we derive the linear stability estimates

$$\|u_\delta - u_\star\| \leq c_1 \delta \,, \tag{I.9}$$

$$\|u_{\alpha,\delta} - u_\star\| \leq c_2 \delta \,, \tag{I.10}$$

for minimizers u_δ, $u_{\alpha,\delta}$ of (I.5) and (I.6), respectively. Note that the subspace $\mathcal{U}_\star$ as well as the constants c_1 and c_2 depend on the operator $\mathbf{K}$ and the signal $u_\star$. A more precise formulation of these statements will be given in Theorem III.8.

A prototypical application we have in mind is deconvolution. Assuming that the convolution operator $\mathbf{K}$ is already in discretized form and invertible, the direct solution $\mathbf{K}^{-1}v_\delta$ of Equation (I.1) provides the linear error estimate $\|\mathbf{K}^{-1}v_\delta - u_\star\| \leq \|\mathbf{K}^{-1}\|\delta$ (which cannot be improved assuming bounded noise). Due to the ill-conditioning of deconvolution, however, the error bound $\|\mathbf{K}^{-1}\|$ is typically enormous and therefore this estimate is useless. On the other hand, the constants c_1 and c_2 in our error estimates for the ℓ^1-programs depend on the norm of $\mathbf{K}^{-1}$ on the possibly small subspace $\mathcal{U}_\star \subset \mathcal{U}$ only. In the case that $u_\star$ is sufficiently sparse and the chosen dictionary $\mathcal{D}$ is sufficiently uncorrelated to the Fourier basis, these constants may be well behaved even for high frequent signals $u_\star$.

E. Relations to Prior Work

Note that the linear error estimates for (I.5) and (I.6) derived in this paper in particular imply the uniqueness of the minimizer of (I.7). In case where $\mathcal{D}$ is an orthonormal basis, uniqueness and linear error estimates have been studied extensively in the compressed sensing context [8], [5], [9], [10], [11], [12], [13], the sparse recovery context [14], [15], [16], [17], [18], [19], as well as in the inverse problems context [3], [4], [20], [2]. Similar results for the redundant case, however, are almost non-existent.

Our main results can be seen as a generalization of the error estimates of [3] from the orthogonal basis case to the case of redundant frames. The architecture of our proofs follows the one given that paper and utilizes error estimates for convex regularization in terms of the Bregman distance. However, the main ingredients for obtaining the actual estimates in terms of the norm distance are quite different. Most notably, in the case of redundant frames one requires injectivity of $\mathbf{K}$ on a finite dimensional subspace spanned the dual frame elements instead of the original frame elements; these subspaces spaces may be completely different. A further (minor) modification is that we introduce an additional positive parameter, that allows balancing two terms arising in the constants in our estimates. This additional freedom can be exploited to obtain even sharper constants in the linear error estimates.

To the best of our knowledge, linear error estimates for the redundant case can only be found in [21], [22], [23]. The estimate derived in [21, Theorem 1.4] requires $\mathcal{D}$ to be a tight fame and is based on a generalization of the restricted isometry property (RIP) from compressed sensing. The error estimate of [23, Theorem 5] generalizes the exact recovery condition of Tropp [19] to the redundant case. Both results and methods are quite different from ours. In [22] error estimates for positively homogeneous regularization functionals have been derived which allow penalties of the form $\|\mathbf{\Phi}^T u\|_1$ with nonorthogonal $\mathbf{\Phi}^T : \mathcal{U} \to \ell^2(\Omega)$. Under a different injectivity assumption [22, Theorem 4.4] yields linear estimates in terms of the regularization functional for the unconstrained program (I.6). In contrast to [22], in the present article we focus on estimates in the Hilbert space norm with explicitly computed constants. Moreover, we consider both the unconstrained program (I.6) as well as the constrained version (I.5).

II. Notation and Preliminaries

Throughout the paper $\mathcal{U}$ and $\mathcal{V}$ denote separable possibly infinite dimensional Hilbert spaces over $\mathbb{R}$ with inner products $\langle \cdot, \cdot \rangle$ and norms $\|\cdot\|$. Moreover, for any at most countable set Ω and any $p \geq 1$ we denote by $\ell^p(\Omega) = \ell^p(\Omega; \mathbb{R})$ the Hilbert space of all real-valued vectors $\{x(\omega) : \omega \in \Omega\}$ such that

$$\|x\|_p^p := \sum_{\omega \in \Omega} |x(\omega)|^p < \infty \,.$$

In the remaining part of this section we shall introduce the notions of frames, subdifferentials and Bregman distance and collect some results required for later purpose.

A. Frames

Recall that the family $\mathcal{D} = \{\phi_\omega : \omega \in \Omega\}$ is a frame of $\mathcal{U}$ if there are constants $0 < a \leq b < \infty$ such that

$$(\forall u \in \mathcal{U}) \qquad a\|u\|^2 \leq \sum_{\omega \in \Omega} |\langle \phi_\omega, u \rangle|^2 \leq b\|u\|^2 \,. \tag{II.1}$$

The smallest and largest numbers a and b, respectively, that satisfy the estimates in (II.1) are called the frame bounds of $\mathcal{D}$. If the upper and lower frame bound coincide, then the frame is said to be tight.

We denote by $\mathbf{\Phi}^T : \mathcal{U} \to \ell^2(\Omega)$ the analysis operator corresponding to $\mathcal{D}$, which takes some signal $u \in \mathcal{U}$ to the coefficient vector $\mathbf{\Phi}^T u := \{\langle \phi_\omega, u \rangle : \omega \in \Omega\}$. Its adjoint is known as synthesis operator and is given by $\mathbf{\Phi}x = \sum_{\omega \in \Omega} \phi_\omega x(\omega)$ for any $x \in \ell^2(\Omega)$. The composition $\mathbf{\Phi}\mathbf{\Phi}^T u = \sum_{\omega \in \Omega} \phi_\omega \langle \phi_\omega, u \rangle$ is called the frame operator. According to the frame property (II.1) it is an invertible linear mapping. In particular, the dual elements $\psi_\omega := (\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\phi_\omega$ are well defined. They again form a frame $\{\psi_\omega : \omega \in \Omega\}$ of $\mathcal{U}$, named the dual frame. The synthesis operator $\mathbf{\Psi}$ corresponding the dual frame is equal to the pseudoinverse $(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}$ of the analysis operator of the original frame. In particular, the signal representation $u = \mathbf{\Psi}\mathbf{\Phi}^T u = \sum_{\omega\in\Omega} \psi_\omega \langle \phi_\omega, u\rangle$ holds for any signal $u \in \mathcal{U}$.

Finally, we say that a signal $u \in \mathcal{U}$ is $\mathcal{D}$-sparse, if $\mathbf{\Phi}^T u$ has at most finitely many nonzero components.

B. Subdifferential and Bregman Distance

An important role in the derivation of error estimates for convex regularization plays the notion of subdifferential.

Definition II.1. Let $\mathrm{R}\colon \mathcal{U} \to \mathbb{R}\cup\{\infty\}$ be a convex functional and let $u_\star$ be any point in $\mathrm{Dom}\,(\mathrm{R}) := \{u \in \mathcal{U} : \mathrm{R}\,(u) < \infty\}$. Then $\eta \in \mathcal{U}$ is called a *subgradient* of R at $u_\star$ if

$$\mathrm{R}\,(u_\star) + \langle \eta, u - u_\star\rangle \leq \mathrm{R}\,(u) \qquad \text{for all } u \in \mathcal{U}\,. \tag{II.2}$$

The set of all subgradients at $u_\star$ is referred to as *subdifferential* of the functional R at $u_\star$ and is denoted by $\partial\mathrm{R}\,(u_\star) \subset \mathcal{U}$.

For the proof of the main results will use estimates for general convex regularization (see Lemma III.1) in terms of the Bregman distance, which is defined as follows.

Definition II.2. Let $\mathrm{R}\colon \mathcal{U} \to \mathbb{R}\cup\{\infty\}$ be a convex functional, let $u, u_\star \in \mathcal{U}$ denote two elements in $\mathcal{U}$ and assume that $\eta \in \partial\mathrm{R}(u_\star)$ is a subgradient of R at $u_\star$. Then the *Bregman distance* between u and $u_\star$ with respect to η is defined by

$$\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star) := \mathrm{R}(u) - \mathrm{R}(u_\star) - \langle \eta, u - u_\star\rangle\ .$$

For a differentiable functional, the subdifferential at some $u_\star$ is single valued and only consist of the usual gradient $\nabla\mathrm{R}\,(u_\star)$. In particular, in the case where $\mathrm{R}\,(u) = 1/2\,\|u\|^2$, one easily sees that $\partial\mathrm{R}(u_\star) = \{u_\star\}$. Moreover, the Bregman distance then reduces to the squared norm distance $\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star) = 1/2\,\|u - u_\star\|^2$. This is, however, not the case for more general functionals, where the Bregman distance may not even be a metric.

For our purpose we will require of the Bregman distance with respect to the ℓ^1-norm, considered as functional on $\ell^2\,(\Omega)$ that takes the value ∞ outside $\ell^1\,(\Omega)$. Note that subgradients of $\|\cdot\|_1$ are elements in the coefficient space $\ell^2\,(\Omega)$ and will be denoted by $\xi = \{\xi(\omega) : \omega \in \Omega\} \in \ell^2\,(\Omega)$ opposed to the notation η for a subgradient in $\mathcal{U}$.

Lemma II.3.

1) *We have $\partial\|\cdot\|_1(x_\star) \neq \emptyset$ if and only if $x_\star \in \ell^2\,(\Omega)$ is a terminating sequence. Moreover, we have*

$$\partial\|\cdot\|_1(x_\star) := \{\xi \in \ell^2(\Omega) : \xi(\omega) \in \mathrm{Sign}(x(\omega))\}\,,$$

where Sign *is defined by* $\mathrm{Sign}(t) = [-1, 1]$ *for* $t = 0$ *and* $\mathrm{Sign}(t) = \{t/|t|\}$ *otherwise.*

2) *The Bregman distance between x and $x_\star$ with respect to $\xi \in \partial\|\cdot\|_1(x_\star)$ is given by*

$$\mathrm{d}_\xi^{\|\cdot\|_1}(x, x_\star) = \sum_{\omega\in\Omega} |x(\omega)| - \xi(\omega)x(\omega)\,.$$

Proof: Both Items are elementary and have been frequently used in ℓ^1-regularization. Since we will repeatedly make use of these facts we include a short proof. For the functional $\|\cdot\|_1$, condition (II.2) reads (with $x_\star, x, \xi \in \ell^2\,(\Omega)$ in place of $u_\star, u, \eta \in \mathcal{U}$)

$$\sum_{\omega\in\Omega} |x_\star(\omega)| + \sum_{\omega\in\Omega} \xi(\omega)\,(x(\omega) - x_\star(\omega)) \leq \sum_{\omega\in\Omega} |x(\omega)|\,,$$

for all $x \in \ell^2\,(\Omega)$. Hence we have $\xi \in \partial\|\cdot\|_1(x_\star)$ if and only if $\xi(\omega) \in \partial|\cdot|\,(x_\star(\omega))$ for all $\omega \in \Omega$. The subdifferential of $|\cdot|$ at some $a_0 \in \mathbb{R}$ is given by $\mathrm{Sign}\,(a_0)$ which yields the claims in Item 1. Now, if $\xi \in \partial\|\cdot\|_1(x_\star)$ one readily verifies $\|x_\star\|_1 = \langle \xi, x_\star\rangle$ which shows Item 2. ■

III. ℓ^1-Analysis Regularization

Throughout the following $\mathcal{D} = \{\phi_\omega : \omega \in \Omega\}$ denotes a frame of $\mathcal{U}$ with frame bounds $a \leq b$, analysis operator $\mathbf{\Phi}^T$, and dual frame $\{\psi_\omega : \omega \in \Omega\}$. Moreover we define the functional $\mathrm{R}\colon \mathcal{U} \to \mathbb{R}\cup\{\infty\}$ by

$$\mathrm{R}\,(u) := \|\mathbf{\Phi}^T u\|_1 = \sum_{\omega\in\Omega} |\langle \phi_\omega, u\rangle|\ . \tag{III.1}$$

Our aim is to derive error estimates for ℓ^1-constrained regularization defined by Equation (I.5) and ℓ^1-Tikhonov regularization defined by Equation (I.6).

A. Error Estimates in the Bregman Distance

From general theory of convex variational regularization we have the following error estimates in terms of the Bregman distance defined in Definition II.2.

Lemma III.1. *Suppose that $u_\star \in \mathcal{U}$ is $\mathcal{D}$-sparse and assume that there exists some element*

$$\eta = \mathbf{K}^T\lambda \in \mathrm{Ran}\left(\mathbf{K}^T\right) \cap \partial\mathrm{R}\,(u_\star)\ . \tag{III.2}$$

Then, for every $\delta > 0$ and every data $v_\delta \in \mathcal{V}$ satisfying $\|\mathbf{K}u_\star - v_\delta\| \leq \delta$ the following assertions hold true.

1) *Every minimizer u_δ of constrained ℓ^1-programm satisfies $\mathrm{d}_\eta^{\mathrm{R}}(u_\delta, u_\star) \leq 2\delta\|\lambda\|$.*
2) *Every minimizer $u_{\alpha,\delta}$ of (I.6) satisfies*

$$\mathrm{d}_\eta^{\mathrm{R}}(u_{\alpha,\delta}, u_\star) \leq \frac{(\delta + \alpha\|\lambda\|/2)^2}{\alpha}\,,$$
$$\|\mathbf{K}u_{\alpha,\delta} - v_\delta\| \leq \delta + \alpha\|\lambda\|\,.$$

Here $\alpha > 0$ is any regularization parameter.

Proof: The estimate in Item 1 has been derived in [24, Theorem 3] and the estimates of Item 2 in [3, Lemma 3.5]. ■

Remark III.2. *Note that Lemma III.1 holds for any convex regularization functional and not only for the ℓ^1-analysis penalty considered in this paper. Condition (III.2) then generalizes the classical source condition $u_\star \in \mathrm{Ran}\left(\mathbf{K}^T\right)$ used for deriving convergence rates for classical Hilbert space regularization (recall that the subdifferential of $1/2\,\|\cdot\|^2$ at some $u_\star$ equals $\{u_\star\}$). However, for non-differential* R, *condition (III.2) does not require $u_\star$ to be smooth but rather imposes* *the existence of a smooth subgradient at $u_\star$. Hence it allows non-smooth signals containing high frequencies.*

Lemma III.3. *Let $u_\star \in \mathcal{U}$ be sparse with respect to $\mathcal{D}$ and let ξ be a subdiffential of $\|\cdot\|_1$ at $\mathbf{\Phi}^T u_\star$.*

1) *We have $\eta := \mathbf{\Phi}\xi \in \partial \mathrm{R}(u_\star) \neq \emptyset$.*
2) *For any $u \in \mathcal{U}$ we have $\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star) = \sum_{\omega\in\Omega} |\langle \phi_\omega, u\rangle| - \xi(\omega)\langle\phi_\omega, u\rangle$.*

Proof: Let ξ be any subgradient of $\|\cdot\|_1$ at $\mathbf{\Phi}^T u_\star$. Since $\mathbf{\Phi}$ is a bounded operator, the element $\eta = \mathbf{\Phi}\xi$ is well defined in $\mathcal{U}$. Moreover, we have

$$\|\mathbf{\Phi}^T u_\star\|_1 + \langle \mathbf{\Phi}\xi, u - u_\star\rangle \\ = \|\mathbf{\Phi}^T u_\star\|_1 + \langle \xi, \mathbf{\Phi}^T u - \mathbf{\Phi}^T u_\star\rangle \leq \|\mathbf{\Phi}^T u\|_1 \,,$$

which shows Item 1. The above identity also shows that $\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star) = \mathrm{d}_\xi^{\|\cdot\|_1}(\mathbf{\Phi}^T u, \mathbf{\Phi}^T u_\star)$. Hence Item 2 follows from the second Item in Lemma II.3. ■

Lemma III.4. *Assume that there exists some subgradient having the form $\eta = \mathbf{K}^T\lambda = \mathbf{\Phi}\xi$ with $\xi \in \partial\|\cdot\|_1(\mathbf{\Phi}^T u_\star)$ and $\lambda \in \mathcal{V}$, and define, for any $t \in (0,1)$, the set*

$$\Omega_\star = \Omega_\star(\xi, t) := \{\omega \in \Omega : |\xi(\omega)| > t\}\,. \tag{III.3}$$

Then the following hold.

1) *$\Omega_\star$ is is a finite set.*
2) *$\max\{|\xi(\omega)| : \omega \notin \Omega_\star\} \leq t$.*
3) *For any $u \in \mathcal{U}$ we have*

$$\sqrt{\sum_{\omega\notin\Omega_\star} |\langle\phi_\omega, u\rangle|^2} \leq \frac{\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star)}{1-t}\,. \tag{III.4}$$

Proof: The first two claims immediately follow from the fact that $\xi \in \ell^2(\Omega)$ and hence must be a sequence converging to zero. Now, Lemma III.3 states that

$$\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star) = \sum_{\omega\in\Omega_\star} |\langle\phi_\omega, u\rangle| - \xi(\omega)\langle\phi_\omega, u\rangle\,.$$

Further, since $t \leq |t|$ for any real number $t \in \mathbb{R}$ and since the ℓ^1-norm $\|\cdot\|_1$ dominates the ℓ^2-norm $\|\cdot\|_2$, we have

$$\begin{aligned} &\sum_{\omega\in\Omega} |\langle\phi_\omega, u\rangle| - \xi(\omega)\langle\phi_\omega, u\rangle \\ &\quad\geq \sum_{\omega\in\Omega} |\langle\phi_\omega, u\rangle| - |\xi(\omega)|\,|\langle\phi_\omega, u\rangle| \\ &\quad\geq (1-t)\sum_{\omega\notin\Omega_\star} |\langle\phi_\omega, u\rangle| \\ &\quad\geq (1-t)\sqrt{\sum_{\omega\notin\Omega_\star} |\langle\phi_\omega, u\rangle|^2}\,. \end{aligned}$$

Combining the last two displayed equations yields Item 3. ■

Remark III.5. *Recall that our aim is the derivation of estimates for the norm-distance $\|u_\delta - u_\star\|$ between the true unknown $u_\star \in \mathcal{U}$ and minimizers u_δ of (I.5) (or minimizers of (I.6)). According to the frame property it is sufficient to derive such estimates for the ℓ^2-norm of the sequence $\{\langle\phi_\omega, u_\delta - u_\star\rangle : \omega \in \Omega\}$.*

Equation (III.4) together with the estimates in Lemma III.1 yield nontrivial estimates for the ℓ^2-norm of $\{\langle\phi_\omega, u_\delta - u_\star\rangle : \omega \notin \Omega_\star\}$. However, for every $t \in (0,1)$, we have $\mathrm{supp}(\mathbf{\Phi}^T u_\star) \subset \Omega_\star$. This implies that (except in the trivial case $u_\star = 0$), the set $\Omega_\star$ is non-empty and that (III.4) does not provide an estimate for the inner products $\langle\phi_\omega, u_\delta - u_\star\rangle$ with $\omega \in \Omega_\star$. In order to establish estimates for these coefficients, and thus finally for the norm distance $\|u_\delta - u_\star\|$, one has to additionally take into account the known estimates for $\|\mathbf{K}u_\delta - \mathbf{K}u_\star\|$ and to make additional assumptions on $u_\star$. In the case that $\mathcal{D}$ is an orthonormal basis, we showed in [3] that the injectivity of $\mathbf{K}$ on $\mathrm{Span}\{\phi_\omega : \omega \in \Omega_\star\}$ is sufficient for that purpose. In the more general frame case considered below we will see that the injectivity on the space spanned by the dual elements ψ_ω (instead of the original ones ϕ_ω) with $\omega \in \Omega_\star$ is the right sufficient assumption (see Assumption III.6 and Theorem III.8 below).

B. Main Error Estimates

In the following we present the main results of this paper. They will be derived under the following assumptions on the interplay between the exact solution $u_\star \in \mathcal{U}$, the frame $\mathcal{D}$, and the operator $\mathbf{K}$; compare with Remark III.5 above.

Assumption III.6.

1) *$\mathbf{K}\colon \mathcal{U} \to \mathcal{V}$ is a bounded linear mapping between the Hilbert spaces $\mathcal{U}$ and $\mathcal{V}$.*
2) *$\mathcal{D} = \{\phi_\omega : \omega \in \Omega\}$ is a frame of $\mathcal{U}$ with dual frame $\{\psi_\omega : \omega \in \Omega\}$.*
3) *The element $u_\star \in \mathcal{U}$ is $\mathcal{D}$-sparse.*
4) *We have $\eta := \mathbf{\Phi}\xi = \mathbf{K}^T\lambda$ for some $\xi \in \partial\|\cdot\|_1(\mathbf{\Phi}^T u_\star)$ and some source element $\lambda \in \mathcal{V}$.*
5) *For some $t \in (0,1)$, the restricted mapping*

$$\mathbf{K}_{\Omega_\star} := \mathbf{K}\big|_{\mathrm{Span}\{\psi_\omega : \omega\in\Omega_\star\}} \tag{III.5}$$

is injective. (Here $\Omega_\star$ is defined by (III.3).)

Remark III.7. *Since the space $\mathrm{Span}\{\psi_\omega : \omega \in \Omega_\star\}$ is finite dimensional (see Lemma III.4), the operator $\mathbf{K}_{\Omega_\star}\colon \mathrm{Span}\{\psi_\omega : \omega \in \Omega_\star\} \to \mathcal{V}$ is boundedly invertible and therefore the operator-norm $\|\mathbf{K}_{\Omega_\star}^{-1}\|$ is well defined and finite.*

We now have the following error estimates for ℓ^1-Tikhonov regularization and constrained regularization with ℓ^1-analysis penalty in terms of norm-distance on $\mathcal{U}$.

Theorem III.8. *Suppose that Assumption III.6 holds, and that $v_\delta \in \mathcal{V}$ are noisy data satisfying $\|\mathbf{K}u_\star - v_\delta\| \leq \delta$. Then the following hold:*

1) *Every minimizer u_δ of (I.5) satisfies $\|u_\delta - u_\star\| \leq c_1\delta$, with*

$$c_1 := 2\left(\|\mathbf{K}_{\Omega_\star}^{-1}\| + \|\lambda\|\,\frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}(1-t)}\right).$$

2) *For the parameter choice $\alpha = C\delta$, every minimizer u_α^δ of (I.6) satisfies $\|u_{\alpha,\delta} - u_\star\| \leq c_2\delta$, with*

$$c_2 := \|\mathbf{K}_{\Omega_\star}^{-1}\|\,(2 + C\|\lambda\|) \\ + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}(1-t)}\,\frac{(1 + C\|\lambda\|/2)^2}{C}\,.$$

Recall that $\Omega_\star$ is defined by (III.3) (and depends on $t \in (0,1)$), that $\mathbf{K}_{\Omega_\star}$ denotes the restriction of $\mathbf{K}$ to the finite dimensional subspace $\operatorname{Span}\{\psi_\omega : \omega \in \Omega_\star\}$, that λ is the source element of Item 4 in Assumption III.6, and that $a > 0$ is the lower frame bound of $\mathcal{D}$.

Proof: The proof will be divided into several Lemmas and will be given in the following Subsection III-C. ∎

C. Proof of Theorem III.8

Let $u_\star$, ξ, η, λ, t and $\mathbf{K}_{\Omega_\star}$ be as in Assumption III.6 and let v_δ be noisy data with $\|\mathbf{K}u_\star - v_\delta\| \le \delta$. Moreover, we denote the standard basis in $\ell^2(\Omega)$ by $\{\mathbf{e}_\omega : \omega \in \Omega\}$. For the following considerations it is convenient to define the orthogonal projectors $\mathbf{\Pi}_{\Omega_\star}, \mathbf{\Pi}_{\Omega_\star^c} : \ell^2(\Omega) \to \ell^2(\Omega)$,

$$\mathbf{\Pi}_{\Omega_\star} x := \sum_{\omega \in \Omega_\star} \mathbf{e}_\omega x(\omega)\,,$$
$$\mathbf{\Pi}_{\Omega_\star^c} := \sum_{\omega \notin \Omega_\star} \mathbf{e}_\omega x(\omega)\,.$$

Finally, we set $\mathbf{P}_{\Omega_\star} := \mathbf{\Psi}\mathbf{\Pi}_{\Omega_\star}\mathbf{\Phi}^T$ and $\mathbf{P}_{\Omega_\star^c} := \mathbf{\Psi}\mathbf{\Pi}_{\Omega_\star^c}\mathbf{\Phi}^T$, where $\mathbf{\Phi}^T$ is the analysis operator corresponding to the frame $\mathcal{D}$ and $\mathbf{\Psi}$ the synthesis operator corresponding to the dual frame $\{\psi_\omega : \omega \in \Omega\}$.

The following relations immediately follow from the above definitions.

Lemma III.9. *For any $u \in \mathcal{U}$, the following hold.*

1) $\mathbf{P}_{\Omega_\star} u = \sum_{\omega \in \Omega_\star} \psi_\omega \langle \phi_\omega, u\rangle$
2) $\mathbf{P}_{\Omega_\star^c} u = \sum_{\omega \notin \Omega_\star} \psi_\omega \langle \phi_\omega, u\rangle$
3) $u = \mathbf{P}_{\Omega_\star} u + \mathbf{P}_{\Omega_\star^c} u$.

Note, however, that the operators $\mathbf{P}_{\Omega_\star}$ and $\mathbf{P}_{\Omega_\star^c}$ are not orthogonal projectors unless $\mathcal{D}$ is an orthogonal basis.

Lemma III.10. *For every $u \in \mathcal{U}$ we have*

$$\|u\| \le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}u\| + \left(1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|\right)\|\mathbf{P}_{\Omega_\star^c} u\|\,.$$

Proof: The triangle inequality and the well-definedness of $\|\mathbf{K}_{\Omega_\star}^{-1}\|$ (see Remark III.7) yield

$$\begin{aligned}\|u\| &\le \|\mathbf{P}_{\Omega_\star} u\| + \|\mathbf{P}_{\Omega_\star^c} u\| \\ &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}_{\Omega_\star}\mathbf{P}_{\Omega_\star} u\| + \|\mathbf{P}_{\Omega_\star^c} u\| \\ &= \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}u - \mathbf{K}\mathbf{P}_{\Omega_\star^c} u\| + \|\mathbf{P}_{\Omega_\star^c} u\| \\ &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}u\| + \left(1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|\right)\|\mathbf{P}_{\Omega_\star^c} u\|\,.\end{aligned}$$

This shows the claim. ∎

Lemma III.11. *For any $u \in \mathcal{U}$ we have*

$$\|\mathbf{P}_{\Omega_\star^c} u\| \le \frac{\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star)}{a^{1/2}\,(1-t)}\,.$$

Proof: From Equation (III.4) it follows that

$$\|\mathbf{\Pi}_{\Omega_\star^c}\mathbf{\Phi}^T u\|_2 \le \frac{1}{1-t}\,\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star)\,.$$

Further, the frame property implies $\|\mathbf{\Psi}\| = a^{-1/2}$. Together with the definition of $\mathbf{P}_{\Omega_\star}$ this yields

$$\begin{aligned}\|\mathbf{P}_{\Omega_\star^c} u\| &= \|\mathbf{\Psi}\mathbf{\Pi}_{\Omega_\star^c}\mathbf{\Phi}^T u\| \\ &\le a^{-1/2}\|\mathbf{\Pi}_{\Omega_\star^c}\mathbf{\Phi}^T u\|_2 \le \frac{\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star)}{a^{1/2}\,(1-t)}\,,\end{aligned}$$

and concludes the proof. ∎

Lemma III.12. *For any $u \in \mathcal{U}$ we have*

$$\begin{aligned}\|u - u_\star\| &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}u - \mathbf{K}u_\star\| \\ &\quad + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\,\mathrm{d}_\eta^{\mathrm{R}}(u, u_\star)\,.\end{aligned}$$

Proof: This follows from Lemmas III.10 and III.11. ∎

Now we are ready to establish Theorem III.8.

Proof of Item 1. Let u_δ denote a minimizer of (I.5). By Item 1 in Lemma III.1 we have $\mathrm{d}_\eta^{\mathrm{R}}(u_\delta, u_\star) \le 2\|\lambda\|\delta$. Moreover since $u_\star$ is an admissible element for (I.5) we have

$$\|\mathbf{K}u_\delta - \mathbf{K}u_\star\| \le \|\mathbf{K}u_\delta - v\| + \|\mathbf{K}u_\star - v\| \le 2\delta\,.$$

Together with Lemma III.12 these estimates imply

$$\begin{aligned}\|u_\delta - u_\star\| &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}u_\delta - \mathbf{K}u_\star\| \\ &\quad + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\,\mathrm{d}_\eta^{\mathrm{R}}(u_\delta, u_\star) \\ &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,2\delta + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\,2\|\lambda\|\delta \\ &= 2\left(\|\mathbf{K}_{\Omega_\star}^{-1}\| + \|\lambda\|\,\frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\right)\delta\,.\end{aligned}$$

Recalling the definition of the constant c_1 yields the final claim $\|u_\delta - u_\star\| \le c_1\delta$.

Proof of Item 2. Let $u_{\alpha,\delta}$ denote a minimizer of the Tikhonov functional in (I.6) with $\alpha = C\delta$. Then by Item 2 in Lemma III.1 and the noise bound $\|\mathbf{K}u_\star - v_\delta\| \le \delta$ we have the estimates

$$\|\mathbf{K}u_{\alpha,\delta} - \mathbf{K}u_\star\| \le 2\delta + \alpha\|\lambda\|\,,$$
$$\mathrm{d}_\eta^{\mathrm{R}}(u_{\alpha,\delta}, u_\star) \le \frac{(\delta + \alpha\|\lambda\|/2)^2}{\alpha}\,.$$

Hence Lemma III.12 and the parameter choice $\alpha = C\delta$ imply that

$$\begin{aligned}\|u_{\alpha,\delta} - u_\star\| &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}u_{\alpha,\delta} - \mathbf{K}u_\star\| \\ &\quad + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\,\mathrm{d}_\eta^{\mathrm{R}}(u_{\alpha,\delta}, u_\star) \\ &\le \|\mathbf{K}_{\Omega_\star}^{-1}\|\,(2\delta + \alpha\|\lambda\|) \\ &\quad + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\,\frac{(\delta + \alpha\|\lambda\|/2)^2}{\alpha} \\ &= \|\mathbf{K}_{\Omega_\star}^{-1}\|\,(2 + C\|\lambda\|)\,\delta \\ &\quad + \frac{1 + \|\mathbf{K}_{\Omega_\star}^{-1}\|\,\|\mathbf{K}\|}{a^{1/2}\,(1-t)}\,\frac{(1 + C\|\lambda\|/2)^2}{C}\,\delta\,.\end{aligned}$$

Recalling the definition of the constant c_2 yields the final claim $\|u_{\alpha,\delta} - u_\star\| \le c_2\delta$ and concludes the proof of Theorem III.8.

IV. Conclusion

In this paper we derived linear error estimates for ℓ^1-analysis regularization of linear ill-posed problems (see Theorem III.8) assuming a sparsity prior for $\mathbf{\Phi}^T u_\star$. Whereas a large amount of such results is available in the orthogonal basis case (see, for example, [5], [10], [15], [17], [12], [3], [19]) our results are among the first for the redundant frame case.

In the case where $\mathcal{D}$ is an orthonormal basis we showed in [3] that Assumption III.6 is not only sufficient but also necessary for the linear error estimates. It would be interesting to know if similar results also hold for redundant systems. Further, it would be interesting to investigate relations between the recovery criteria $\mathrm{RC}(I)$ of [23], which allows linear error estimates for ℓ^1-analysis minimization (at least in the finite dimensional setting), and our Assumption III.6.

References

[1] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, ser. Mathematics and its Applications. Dordrecht: Kluwer Academic Publishers Group, 1996, vol. 375.

[2] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational methods in imaging*, ser. Applied Mathematical Sciences. New York: Springer, 2009, vol. 167.

[3] M. Grasmair, M. Haltmeier, and O. Scherzer, "Necessary and sufficient conditions for linear convergence of ℓ^1-regularization," *Comm. Pure Appl. Math.*, vol. 64, no. 2, pp. 161–182, 2011. [Online]. Available: http://dx.doi.org/10.1002/cpa.20350

[4] ——, "Sparse regularization with l^q penalty term," *Inverse Probl.*, vol. 24, no. 5, pp. 055 020, 13, 2008. [Online]. Available: http://dx.doi.org/10.1088/0266-5611/24/5/055020

[5] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006. [Online]. Available: http://dx.doi.org/10.1002/cpa.20124

[6] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Probl.*, vol. 23, no. 3, p. 947, 2007. [Online]. Available: http://dx.doi.org/10.1088/0266-5611/23/3/007

[7] S. Nama, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal., in press*, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.acha.2012.03.006

[8] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2006.885507

[9] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007. [Online]. Available: http://dx.doi.org/10.1109/LSP.2007.898300

[10] D. L. Donoho, "For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," *Comm. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2006. [Online]. Available: http://dx.doi.org/10.1002/cpa.20131

[11] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011. [Online]. Available: http://dx.doi.org/10.1109/TSP.2011.2161982

[12] S. Foucart., "A note on guaranteed sparse recovery via ℓ^1-minimization," *Appl. Comput. Harmon. Anal.*, vol. 29, no. 1, pp. 97–103, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.acha.2009.10.004

[13] T. T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ^1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, 2009. [Online]. Available: http://dx.doi.org/10.1109/TIT.2009.2021377

[14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003. [Online]. Available: http://dx.doi.org/10.1073/pnas.0437847100

[15] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2005.860430

[16] J. J. Fuchs, "On sparse representations in arbitrary redundant basis," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004. [Online]. Available: http://dx.doi.org/10.1109/TIT.2004.828141

[17] ——, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, 2005. [Online]. Available: http://dx.doi.org/10.1109/TIT.2005.855614

[18] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003. [Online]. Available: http://dx.doi.org/10.1109/TIT.2003.820031

[19] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2005.864420

[20] D. A. Lorenz, S. Schiffler, and D. Trede, "Beyond convergence rates: exact recovery with the Tikhonov regularization with sparsity constraints," *Inverse Probl.*, vol. 27, no. 8, p. 085009, 2011. [Online]. Available: http://dx.doi.org/10.1088/0266-5611/27/8/085009

[21] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.acha.2010.10.002

[22] M. Grasmair, "Linear convergence rates for Tikhonov regularization with positively homogeneous functionals," *Inverse Probl.*, vol. 27, no. 7, p. 075014, 2011. [Online]. Available: http://dx.doi.org/10.1088/0266-5611/27/7/075014

[23] S. Vaiter, G. Peyrè, C. Dossal, and J. Fadili, "Robust sparse analysis regularization," arXiv:1109.6222v3 [cs.IT], 2012. [Online]. Available: http://arxiv.org/abs/1109.6222

[24] M. Burger and S. Osher, "Convergence rates of convex variational regularization," *Inverse Probl.*, vol. 20, no. 5, pp. 1411–1421, 2004. [Online]. Available: http://dx.doi.org/10.1088/0266-5611/20/5/005