

# Sparsity and Compressed Sensing in Inverse Problems

Evelyn Herrholz and Dirk Lorenz and Gerd Teschke and Dennis Trede

**Abstract** This chapter is concerned with two important topics in the context of sparse recovery in inverse and ill-posed problems. In first part we elaborate conditions for exact recovery. In particular, we describe how both  $\ell^1$ -minimization and matching pursuit methods can be used to regularize ill-posed problems and moreover, state conditions which guarantee exact recovery of the support in the sparse case. The focus of the second part is on the incomplete data scenario. We discuss extensions of compressed sensing for specific infinite dimensional ill-posed measurement regimes. We are able to establish recovery error estimates when adequately relating the isometry constant of the sensing operator, the ill-posedness of the underlying model operator and the regularization parameter. Finally, we very briefly sketch how projected steepest descent iterations can be applied to retrieve the sparse solution.

---

Dr. Evelyn Herrholz

Hochschule Neubrandenburg, Institute for Computational Mathematics in Science and Technology, Brodaer Str. 2, 17033 Neubrandenburg, e-mail: evelyn@herrholz.de

Prof. Dr. Dirk Lorenz,

Technische Universität Braunschweig, Institut für Analysis und Algebra, 38092 Braunschweig, e-mail: d.lorenz@tu-braunschweig.de

Prof. Dr. Gerd Teschke

Hochschule Neubrandenburg, Institute for Computational Mathematics in Science and Technology, Brodaer Str. 2, 17033 Neubrandenburg, e-mail: teschke@hs-nb.de

Dr. Dennis Trede

Universität Bremen, Zentrum für Technomathematik, 28334 Bremen, e-mail: trede@math.uni-bremen.de

## 1 Introduction

Many applications in science and engineering require the solution of an operator equation  $Kx = y$ . Often only noisy data  $y^\delta$  with  $\|y^\delta - y\| \leq \delta$  are available, and if the problem is ill-posed, regularization methods have to be applied. During the last three decades, the theory of regularization methods for treating linear problems in a Hilbert space framework has been well developed, see, e.g., [23, 29, 30, 39]. Influenced by the huge impact of sparse signal representations and the practical feasibility of advanced sparse recovery algorithms, the combination of sparse signal recovery and inverse problems emerged in the last decade as a new growing area. Currently, there exist a great variety of sparse recovery algorithms for inverse problems (linear as well as for nonlinear operator equations) within this context, see, e.g., [5, 6, 7, 14, 15, 16, 25, 26, 41, 44, 45]. These recovery algorithms are successful for many applications and have led to breakthroughs in many fields. However, the feasibility is usually limited to problems for which the data are complete and where the problem is of moderate dimension. For really large-scale problems or problems with incomplete data, these algorithms are not well-suited and often far off exact recovery or fail completely.

Within this chapter we focus on two neighboring questions arising in sparse recovery of solutions of inverse problems. The first is concerned with *exact recovery conditions in the complete data scenario*, and the second is concerned with *sparse recovery in the compressively sensed data scenario*.

*Exact recovery.* The two most widely used recovery methods, namely  $\ell^1$ -minimization and matching pursuit methods, can be related to two classical methods for regularization of ill-posed problems:  $\ell^1$ -minimization is a special case of variational regularization in which the operator equation  $Kx = y$  is replaced by a well-posed minimization problem with a sparsity constraint. Matching pursuit methods are related to iterative regularization methods in which one uses an iterative method to solve the operator equation and uses a stopping criterion to prevent noise amplification. We describe how both  $\ell^1$ -minimization and matching pursuit methods can be used to regularize ill-posed problems and moreover, state conditions which guarantee exact recovery of the support in the sparse case.

*Compressive sensing.* For the incomplete data situation, the mathematical technology called compressive sensing, which turned out to be quite successful in sparse signal recovery, was established several years ago by D. Donoho, see [18]. A major breakthrough was achieved when it was proven that it is possible to reconstruct a signal from very few measurements under certain conditions on the signal and the measurement model, see [8, 9, 10, 19, 20, 18, 24, 42]. In [12] it was shown that if the sensing operator satisfies the restricted isometry property the solution can be reconstructed exactly by minimization of an  $\ell_1$  constrained problem, provided that the solution is sparse enough. Classical formulations of compressed sensing are finite dimensional. Quite recently, continuous formulations have appeared, see [1] (full continuous sensing model) and see, e.g., [22, 33, 38] (problem of analog-to-digital conversion). Within this chapter we summarize extensions of the infinite dimensional model in [22] to the case of compressively sampling ill-posed problems and

provide iterative sparse recovery principles and corresponding error estimates, for detailed discussions see [31]. Further extensions towards generalized and compressive sampling also on the context of ill-posed problems can be found in [3], [2], and [4].

## 2 Exact recovery for ill-posed problems

In this section we describe how both  $\ell^1$ -minimization and matching pursuit methods can be used to regularize ill-posed problems and moreover, state condition which guarantee exact recovery of the support in the sparse case.

### 2.1 Orthogonal matching pursuit

In a Banach space  $X$ , we assume that we have a given *dictionary* of unit-normed atoms  $(e_i) = \mathcal{E}$ . We assume that the solutions to an operator equation  $Kx = y$  (with  $K : X \rightarrow Y$  bounded, injective and linear and  $Y$  a Hilbert space) can be expressed sparsely in  $\mathcal{E}$ , i.e. that

$$x = \sum_{i \in \mathbb{Z}} \alpha_i e_i \quad \text{with} \quad \alpha_i \in \mathbb{R}, \quad \|\alpha\|_{\ell^0} =: N < \infty.$$

Now assume that instead of  $y = Kx$  we are given a noisy measurement  $y^\varepsilon$  with  $\|y - y^\varepsilon\| \leq \varepsilon$  and aim to recover a good approximation of  $x$  from the measurement  $y^\varepsilon$ .

In the following we denote with  $I$  the support of the coefficient vector  $\alpha$ , i.e.  $I = \{i \in \mathbb{Z} \mid \alpha_i \neq 0\}$ . For any subset  $J \subset \mathbb{Z}$  we denote  $\mathcal{E}(J) := \{e_i \mid i \in J\}$ .

The above setting is of practical relevance, e.g. in mass spectrometry [32] where the signal is modeled as a sum of Dirac peaks (so-called impulse trains)  $x = \sum_{i \in \mathbb{Z}} \alpha_i \delta(\cdot - t_i)$ . Another example can be found in digital droplet holography, cf. [43], where images arise as superposition of characteristic functions of balls with different centers  $t_i$  and radii  $r_j$ ,  $x = \sum_{i,j \in \mathbb{Z}} \alpha_{i,j} \chi_{B_{r_j}}(\cdot - t_i)$ .

In this section we approach the problem “ $Kx = y^\varepsilon$ ” by iteratively including more and more atoms in the representation of  $x$  in a “greedy” fashion—an algorithmic idea which is also known under the name *matching pursuit*. We define another normed dictionary

$$\mathcal{D} := \{d_i\}_{i \in \mathbb{Z}} := \left\{ \frac{Ke_i}{\|Ke_i\|} \right\}_{i \in \mathbb{Z}}.$$

Note that  $\mathcal{D}$  is well defined by the injectivity of  $K$ . In any step of our iterative method we select that atom from the dictionary  $\mathcal{D}$  which is correlates most with the residual (hence the name “greedy” method). To stabilize the solution of “ $Kx =$

$y^\varepsilon$ ” the iteration has to be stopped early enough. We only investigate the so-called *orthogonal matching pursuit* (OMP), first proposed in the signal processing context by Davis et al. in [36] and Pati et al. in [40] as an improvement upon the matching pursuit algorithm [37]. The algorithm is stated in Figure 1.

**Fig. 1** Orthogonal Matching Pursuit

1. Initialize  $k = 0, I^0 = \emptyset, r^0 = y^\varepsilon, \hat{x}^0 = 0$
2. While  $\|r_k\| > \varepsilon$  do:
  - a. Increase  $k$  and select an atom by  $i_k \in \operatorname{argsup} \{|\langle r^{k-1}, d_i \rangle| \mid d_i \in \mathcal{D}\}$ ,
  - b. Set  $I^k = I^{k-1} \cup \{i_k\}$  and project  $x$  onto  $\operatorname{span} \mathcal{E}(I^k)$ , i.e.  $\hat{x}^k = \operatorname{argmin} \{\|y^\varepsilon - K\hat{u}\|^2 \mid \hat{u} \in \operatorname{span} \mathcal{E}(I^k)\}$ ,
  - c. Set  $r^k := y^\varepsilon - K\hat{x}^k$ .

Necessary and sufficient conditions for exact support recovery by OMP are given in [46]. Next, we list this result in the language of infinite-dimensional inverse problems. We define the linear continuous synthesis operator for the dictionary  $\mathcal{D}$  via  $D : \ell^1 \rightarrow Y, D\beta = \sum_{i \in \mathbb{Z}} \beta_i d_i = \sum_{i \in \mathbb{Z}} \beta_i \frac{K e_i}{\|K e_i\|}$ . Furthermore, for  $J \subset \mathbb{Z}$  we denote with  $P_J : \ell^1 \rightarrow \ell^1$  the projection onto  $J$  and with  $A^\dagger$  the pseudoinverse of an operator  $A$ . With this notation we state the following theorem.

**Theorem 1 (Tropp [46]).** *Let  $\alpha \in \ell^0$  with  $\operatorname{supp} \alpha = I, x = \sum_{i \in \mathbb{Z}} \alpha_i e_i$  be the source and  $y = Kx$  the measured signal. If the operator  $K : X \rightarrow Y$  and the dictionary  $\mathcal{E} = \{e_i\}_{i \in \mathbb{Z}}$  fulfill the Exact Recovery Condition (ERC)*

$$\sup_{d \in \mathcal{D}(I^c)} \|(DP_I)^\dagger d\|_{\ell^1} < 1, \quad (1)$$

*then OMP with its parameter  $\varepsilon$  set to 0 recovers  $\alpha$  exactly.*

The necessity of the condition (1) is shown in [46], by constructing a signal such that for  $\geq 1$  in (1), OMP fails to recover it.

A weaker sufficient condition is derived by Dossal and Mallat [21] and Gribonval and Nielsen [28] and this condition only depends on inner products of the dictionary atoms of  $\mathcal{D}(I)$  and  $\mathcal{D}(I^c)$  only and hence, is simpler to evaluate (although the condition is not necessary).

**Proposition 1 (Dossal and Mallat [21], Gribonval and Nielsen [28]).** *Let  $\alpha \in \ell^0$  with  $\operatorname{supp} \alpha = I$ . If the operator  $K : X \rightarrow Y$  and the dictionary  $\mathcal{E} = \{e_i\}_{i \in \mathbb{Z}}$  fulfill the Neumann ERC*

$$\sup_{i \in I} \sum_{j \in I, j \neq i} |\langle d_i, d_j \rangle| + \sup_{i \in I^c} \sum_{j \in I} |\langle d_i, d_j \rangle| < 1, \quad (2)$$

*then OMP with its parameter  $\varepsilon$  set to 0 recovers  $\alpha$ .*

The transfer to noisy signals  $y^\varepsilon = y + \eta = Kx + \eta$  with  $\|v - v^\varepsilon\| = \|\eta\| \leq \varepsilon$  (where OMP has to stop as soon as  $\varepsilon \geq \|r^k\|$ ) is contained in the following theorem from [17].

**Theorem 2 (ERC in the Presence of Noise).** *Let  $\alpha \in \ell^0$  with  $\text{supp } \alpha = I$ . Let  $x = \sum_{i \in \mathbb{Z}} \alpha_i e_i$  be the source and  $y^\varepsilon = Kx + \eta$  the noisy data with noise level  $\|\eta\| \leq \varepsilon$  and noise-to-signal-ratio*

$$r_{\varepsilon/\alpha} := \frac{\sup_{i \in \mathbb{Z}} |\langle \eta, d_i \rangle|}{\min_{i \in I} |\alpha_i| \|Ke_i\|}.$$

*If the operator  $K$  and the dictionary  $\mathcal{E}$  fulfill the Exact Recovery Condition in Presence of Noise ( $\varepsilon$ ERC)*

$$\sup_{d \in \mathcal{D}(I^c)} \|(DP_I)^\dagger d\|_{\ell^1} < 1 - 2r_{\varepsilon/\alpha} \frac{1}{1 - \sup_{i \in I} \sum_{j \in I, j \neq i} |\langle d_i, d_j \rangle|}, \quad (3)$$

*and  $\sup_{i \in I} \sum_{j \in I, j \neq i} |\langle d_i, d_j \rangle| < 1$ , then OMP recovers the support  $I$  of  $\alpha$  exactly.*

To ensure the  $\varepsilon$ ERC (3) one has necessarily for the noise-to-signal-ratio  $r_{\varepsilon/\alpha} < 1/2$ . A rough upper bound for  $\sup_{i \in \mathbb{Z}} |\langle \eta, d_i \rangle|$  is  $\varepsilon$  and hence, one may use  $r_{\varepsilon/\alpha} \leq \varepsilon / (\min_{i \in I} |\alpha_i| \|Ke_i\|)$ .

Similarly to the result of Dossal and Mallat, one can give a weaker sufficient recovery condition that depends on inner products of the dictionary atoms. It is proved analogously to proposition 1 (see [17]).

**Proposition 2 (Neumann ERC in the Presence of Noise).** *If the operator  $K$  and the dictionary  $\mathcal{E}$  fulfill the Neumann  $\varepsilon$ ERC*

$$\sup_{i \in I} \sum_{j \in I, j \neq i} |\langle d_i, d_j \rangle| + \sup_{i \in I^c} \sum_{j \in I} |\langle d_i, d_j \rangle| < 1 - 2r_{\varepsilon/\alpha}, \quad (4)$$

*then OMP recovers the support  $I$  of  $\alpha$  exactly.*

Theorem 2 and proposition 2 ensure that the correct support  $I$  is identified and the following proposition additionally shows that the reconstruction error is of the order of the noise level.

**Proposition 3 (Error bounds for OMP in presence of noise).** *If the  $\varepsilon$ ERC is fulfilled then there exists a constant  $C > 0$  such that for the approximative solution  $\hat{\alpha}$  determined by OMP it holds that  $\|\hat{\alpha} - \alpha\|_{\ell^1} \leq C\varepsilon$ .*

The proof can also be found in [17]

## 2.2 $\ell^1$ -minimization

In  $\ell^1$ -minimization one promotes sparsity of the approximate solution of  $Kx = y^\varepsilon$  by a sparsity constraint. In this section we assume that  $x$  itself is the object which is sparse, i.e.  $x \in \ell^2$  with  $|\text{supp}x| = N < \infty$ . A typical sparsity constraint is given by the  $\ell^1$ -norm and hence, we investigate the minimization problem

$$\min_x \left\{ T_\lambda(x) = \frac{1}{2} \|Kx - y^\varepsilon\|^2 + \lambda \|x\|_{\ell^1} \right\}.$$

This method is also called Basis Pursuit Denoising [13].

In [34, 27] it has been shown that  $\ell^1$  minimization is indeed a regularization method and also an error estimate have been derived. A central ingredient is the so-called Finite Basis Injectivity property (FBI-property) of the operator  $K$  introduced in [7]. An operator  $K$  has the FBI property if for all finite subsets  $J \subset \mathbb{Z}$  the operator  $K$  restricted to  $\text{span}\{e_i | i \in J\}$  is injective, (in other words, for all  $x, z \in \ell^2$  with  $Kx = Kz$  and  $x_k = z_k = 0$ , for all  $k \notin J$ , it follows that  $x = z$ ). Note that the FBI property can be seen as a variant of the restricted isometry property (introduced in the next section).

**Theorem 3 (Error estimate).** *Let  $K$  possess the FBI property,  $x$  be sparse with  $\text{supp}x = I$  be a minimum- $\|\cdot\|_{\ell^1}$  solution of  $Kx = y$ , and  $\|y - y^\varepsilon\| \leq \varepsilon$ . Let the following source condition (SC) be fulfilled:*

$$\text{there exists } w \in Y \text{ such that } K^*w = \xi \in \text{Sign}(x). \quad (5)$$

Moreover, let  $\theta = \sup\{|\xi_k| \mid |\xi_k| < 1\}$  and  $c > 0$  such that for all  $z \in \ell^2$  with  $\text{supp}(z) \subset I$  it holds  $\|Ku\| \geq c\|u\|$ . Then for the minimizers  $x^{\lambda, \varepsilon}$  of  $T_\lambda$  it holds

$$\|x^{\lambda, \varepsilon} - x\|_{\ell^1} \leq \frac{\|K\| + 1}{1 - \theta} \frac{\varepsilon^2}{\lambda} + \left( \frac{1}{c} + \|w\| \frac{\|K\| + 1}{1 - \theta} \right) (\lambda + \varepsilon). \quad (6)$$

Especially, with  $\lambda \asymp \varepsilon$  it holds

$$\|x^{\lambda, \varepsilon} - x\|_{\ell^1} = \mathcal{O}(\varepsilon). \quad (7)$$

In addition to the above error estimate one can give an a priori parameter rule which ensures that the unknown support of the sparse solution  $x \in \ell^0$  is recovered exactly (cf. [35]).

**Theorem 4 (Lower bound on  $\alpha$ ).** *Let  $x \in \ell^0$ ,  $\text{supp}(x) = I$ , and  $y^\varepsilon = Kx + \eta$  the noisy data. Assume that  $K$  is bounded and possesses the FBI property. If the following condition holds,*

$$\sup_{i \in I^c} \|(KP_I)^\dagger Ke_i\|_{\ell^1} < 1, \quad (8)$$

then the parameter rule

$$\alpha > \frac{1 + \sup_{i \in \mathcal{I}} \|(KP_I)^\dagger Ke_i\|_{\ell^1}}{1 - \sup_{i \in \mathcal{I}} \|(KP_I)^\dagger Ke_i\|_{\ell^1}} \sup_{i \in \mathbb{Z}} |\langle \eta, Ke_i \rangle| \quad (9)$$

ensures that the support of  $x^{\lambda, \varepsilon}$  is contained in  $I$ .

Theorem 4 gives a lower bound on the regularization parameter  $\lambda$  to ensure  $\text{supp}(x^{\lambda, \varepsilon}) \subset \text{supp}(x)$ . To even guarantee  $\text{supp}(x^{\lambda, \varepsilon}) = \text{supp}(x)$  we need an additional upper bound for  $\lambda$ . The following theorem from [35] leads to that purpose.

**Theorem 5 (Error estimate).** *Let the assumptions of theorem 4 hold and choose  $\lambda$  according to (9). Then the following error estimate is valid:*

$$\|x - u^{\lambda, \varepsilon}\|_{\ell^\infty} \leq (\lambda + \sup_{i \in \mathbb{Z}} |\langle \eta, Ke_i \rangle|) \|(P_I K^* K P_I)^{-1}\|_{\ell^1, \ell^1}. \quad (10)$$

*Remark 1.* Due to the error estimate (10) we achieve a linear convergence rate measured in the  $\ell^\infty$  norm. In finite dimensions all  $\ell^p$  norms are equivalent, hence we also get an estimate for the  $\ell^1$  error:

$$\|x - x^{\lambda, \varepsilon}\|_{\ell^1} \leq (\lambda + \varepsilon \|K\|) |I| \|(P_I K^* K P_I)^{-1}\|_{\ell^1, \ell^1}.$$

Compared to the estimate (6) from theorem 3, the quantities  $\theta$  and  $\|w\|$  are not present anymore. The role of  $1/c$  is now played by  $\|(P_I K^* K P_I)^{-1}\|_{\ell^1, \ell^1}$ . However, if upper bounds on  $I$  or on its size (together with structural information on  $K$ ) is available, the estimate can give a-priori checkable error estimates.

**Theorem 6 (Exact recovery condition in the presence of noise).** *Let  $x \in \ell^0$  with  $\text{supp}(x) = I$  and  $y^\varepsilon = Kx + \eta$  the noisy data with noise-to-signal ratio*

$$r_{\eta/u} := \frac{\sup_{i \in \mathbb{Z}} |\langle \eta, Ke_i \rangle|}{\min_{i \in I} |x_i|}.$$

*Assume that the operator  $K$  is bounded and possesses the FBI property. Then the exact recovery condition in the presence of noise ( $\varepsilon$ ERC)*

$$\sup_{i \in \mathcal{I}^c} \|(KP_I)^\dagger Ke_i\|_{\ell^1} < 1 - 2r_{\eta/u} \|(P_I K^* K P_I)^{-1}\|_{\ell^1, \ell^1} \quad (11)$$

*ensures that there is a suitable regularization parameter  $\lambda$ ,*

$$\frac{1 + \sup_{i \in \mathcal{I}^c} \|(KP_I)^\dagger Ke_i\|_{\ell^1}}{1 - \sup_{i \in \mathcal{I}^c} \|(KP_I)^\dagger Ke_i\|_{\ell^1}} \sup_{i \in \mathbb{Z}} |\langle \eta, Ke_i \rangle| < \lambda \quad (12)$$

$$\lambda < \frac{\min_{i \in I} |u_i^\diamond|}{\|(P_I K^* K P_I)^{-1}\|_{\ell^1, \ell^1}} - \sup_{i \in \mathbb{Z}} |\langle \eta, Ke_i \rangle|,$$

*which provides exact recovery of  $I$ .*

### 3 Compressive sensing principles for ill-posed problems

Within this section we combine the concepts of compressive sensing and sparse recovery for solving inverse and ill-posed problems. To establish an adequate measurement model, we adapt an infinite dimensional compressed sensing setup that was invented in [22]. As the main result we provide recovery accuracy estimates for the computed sparse approximations in the language of [11] but now for the solution of the underlying inverse problem. One essential difference to the classical compressed sensing framework is the incorporation of joint sparsity measures allowing the treatment of infinite dimensional reconstruction spaces. Moreover, to tackle ill-posed operator equations we rely on constrained optimization formulations that are very close to elastic net type optimizations.

#### 3.1 Compressive sensing model and classical results

Within this section we provide the standard reconstruction space, the compressive sensing model and repeat classical recovery results for finite-dimensional problems that can be established thanks to the restricted isometry property of the underlying sensing matrix.

Let  $X$  be a separable Hilbert space and  $X_m \subset X$  the (possibly infinite dimensional) reconstruction space defined by

$$X_m = \left\{ x \in X, x = \sum_{\ell=1}^m \sum_{\lambda \in \Lambda} d_{\ell,\lambda} a_{\ell,\lambda}, d \in (\ell_2(\Lambda))^m \right\},$$

where we assume that  $\Lambda$  is a countable index set and  $\Phi_a = \{a_{\ell,\lambda}, \ell = 1, \dots, m, \lambda \in \Lambda\}$  forms a frame for  $X_m$  with frame bounds  $0 < C_{\Phi_a} \leq C^{\Phi_a} < \infty$ . Note that the reconstruction space  $X_m$  is a subspace of  $X$  with possibly large  $m$ . Typically we consider functions of the form  $a_{\ell,\lambda} = a_\ell(\cdot - \lambda \mathcal{T})$ , for some  $\mathcal{T} > 0$ . With respect to  $\Phi_a$  we define the map  $F_a : X_m \rightarrow (\ell_2(\Lambda))^m$  through  $x \mapsto F_a x = (\{\langle x, a_{1,\lambda} \rangle\}_{\lambda \in \Lambda}, \dots, \{\langle x, a_{m,\lambda} \rangle\}_{\lambda \in \Lambda})^T$ .  $F_a$  is the analysis operator and its adjoint, given by  $F_a^* : (\ell_2(\Lambda))^m \rightarrow X_m$  through  $d \mapsto F_a^* d = \sum_{\ell=1}^m \sum_{\lambda \in \Lambda} d_{\ell,\lambda} a_{\ell,\lambda}$ , is the so-called synthesis operator. Since  $\Phi_a$  forms a frame, each  $x \in X_m$  can be reconstructed from its moments  $F_a x$  through  $(F_a^* F_a)^{-1} F_a^*$ . A special choice of analysis/sampling functions might relax the situation a bit. Assume we have another family of sampling functions  $\Phi_v$  at our disposal fulfilling  $F_v F_a^* = I$ , then it follows with  $x = F_a^* d$

$$y = F_v x = \begin{pmatrix} \{\langle x, v_{1,\lambda} \rangle\}_{\lambda \in \Lambda} \\ \vdots \\ \{\langle x, v_{m,\lambda} \rangle\}_{\lambda \in \Lambda} \end{pmatrix} = \begin{pmatrix} \{\langle F_a^* d, v_{1,\lambda} \rangle\}_{\lambda \in \Lambda} \\ \vdots \\ \{\langle F_a^* d, v_{m,\lambda} \rangle\}_{\lambda \in \Lambda} \end{pmatrix} = F_v F_a^* d = d, \quad (13)$$



i.e. the sensed values  $y$  equal  $d$  and therefore  $x = F_a^* F_v x$ . The condition  $F_v F_a^* = I$  means nothing else than  $\langle a_{\ell, \lambda}, v_{\ell', \lambda'} \rangle = \delta_{\lambda \lambda'} \delta_{\ell \ell'}$  for all  $\lambda, \lambda' \in \Lambda$  and  $\ell, \ell' = 1, \dots, m$ , i.e.  $\Phi_v$  and  $\Phi_a$  are biorthogonal to each other.

As we focus on reconstructing functions (or solutions of operator equations)  $x$  that have a sparse series expansion  $x = F_a^* d$  with respect to  $\Phi_a$ , i.e. the series expansion of  $x$  has only a very small number of non-vanishing coefficients  $d_{\ell, \lambda}$ , or that  $x$  is compressible (meaning that  $x$  can be well-approximated by a sparse series expansion), the theory of compressed sensing suggests to sample  $x$  at much lower rate as done in the classical setting mentioned above (there it was  $m/\mathcal{T}$ ) while ensuring exact recovery of  $x$  (or recovery with overwhelming probability). The compressive sampling idea applied to the sensing situation (13) goes now as follows. Assume we are given a sensing matrix  $A \in \mathbb{R}^{p \times m}$  with  $p \ll m$ . Then we construct  $p$  species of sampling functions through

$$\begin{pmatrix} s_{1, \lambda} \\ \vdots \\ s_{p, \lambda} \end{pmatrix} = A \begin{pmatrix} v_{1, \lambda} \\ \vdots \\ v_{m, \lambda} \end{pmatrix} \quad \text{for all } \lambda \in \Lambda. \quad (14)$$

As a simple consequence of (14), the following lemma holds true.

**Lemma 1.** *Assume for all  $\lambda \in \Lambda$  the sampling functions  $s_{1, \lambda}, \dots, s_{p, \lambda}$  are chosen as in (14) and let  $y$  denote the exactly sensed data. If  $\Phi_a$  and  $\Phi_v$  are biorthogonal to each other, then  $y = Ad$ .*

Let  $d_\lambda$  denote the  $m$ -dimensional vector  $(d_{1, \lambda}, \dots, d_{m, \lambda})^T$  and  $y_\lambda$  the  $p$ -dimensional vector  $(y_{1, \lambda}, \dots, y_{p, \lambda})^T$ , then Lemma 1 states that for each  $\lambda \in \Lambda$  the measurement vectors are given by  $y_\lambda = Ad_\lambda$ . It has been shown in [12], that for each individual  $\lambda \in \Lambda$  the solution  $d_\lambda^*$  to

$$\min_{d_\lambda \in \mathbb{R}^m} \|d_\lambda\|_{\ell_1} \quad \text{subject to } y_\lambda = Ad_\lambda, \quad (15)$$

recovers  $d_\lambda$  exactly provided that  $d_\lambda$  is sufficiently sparse and the matrix  $A$  obeys a condition known as the *restricted isometry property*.

**Definition 1 (restricted isometry property).** For each integer  $k = 1, 2, \dots$ , define the isometry constant  $\delta_k$  of a sensing matrix  $A$  as the smallest number such that

$$(1 - \delta_k) \|x\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta_k) \|x\|_{\ell_2}^2 \quad (16)$$

holds for all  $k$ -sparse vectors  $x$ . A vector is said to be  $k$ -sparse if it has at most  $k$  non-vanishing entries.

**Theorem 7 (noiseless recovery, Candès [11]).** *Assume  $\delta_{2k} < \sqrt{2} - 1$ . Then for each  $\lambda \in \Lambda$  the solution  $d_\lambda^*$  to (15) obeys*

$$\|d_\lambda^* - d_\lambda\|_{\ell_1} \leq C_0 \|d_\lambda^k - d_\lambda\|_{\ell_1} \quad (17)$$

$$\|d_\lambda^* - d_\lambda\|_{\ell_2} \leq C_0 k^{-1/2} \|d_\lambda^k - d_\lambda\|_{\ell_1} \quad (18)$$

for some constant  $C_0$  (that can be explicitly computed) and  $d_\lambda^k$  denoting the best  $k$ -term approximation. If  $d_\lambda$  is  $k$ -sparse, the recovery is exact.

This result can be extended to the more realistic scenario in which the measurements are contaminated by noise. Then we have to solve

$$\min_{d_\lambda \in \mathbb{R}^m} \|d_\lambda\|_{\ell_1} \quad \text{subject to} \quad \|y_\lambda^\delta - Ad_\lambda\|_{\ell_2} \leq \delta. \quad (19)$$

**Theorem 8 (noisy recovery, Candès [11]).** Assume  $\delta_{2k} < \sqrt{2} - 1$  and  $\|y_\lambda^\delta - y_\lambda\|_{\ell_2} \leq \delta$ . Then for each  $\lambda \in \Lambda$  the solution  $d_\lambda^*$  to (19) obeys

$$\|d_\lambda^* - d_\lambda\|_{\ell_2} \leq C_0 k^{-1/2} \|d_\lambda^k - d_\lambda\|_{\ell_1} + C_1 \delta \quad (20)$$

with the same constant  $C_0$  as before and some  $C_1$  (that can be explicitly computed).

### 3.2 Infinite dimensional regime and joint sparsity measures

In the previous subsection we have summarized results that apply for all individual sensing scenarios, i.e. that hold true for all individual  $\lambda \in \Lambda$ . But as the index set  $\Lambda$  is possibly of infinite cardinality, we are faced with the problem of recovering infinitely many unknown vectors  $d_\lambda$  for which the (essential) support can be different. Therefore, the determination of  $d$  by solving for each  $\lambda$  an individual optimization problem is numerically not feasible.

For a simultaneous treatment of all individual optimization problems, we have to restrict the set of all possible solutions  $d_\lambda$ . One quite natural restriction is that all  $d_\lambda$  share a joint sparsity pattern. Introducing support sets  $\mathcal{S} \subset \{1, \dots, m\}$ , the reconstruction space is given through

$$X_k = \left\{ x \in X, x = \sum_{\ell \in \mathcal{S}, |\mathcal{S}|=k} \sum_{\lambda \in \Lambda} d_{\ell, \lambda} a_{\ell, \lambda}, d \in (\ell_2(\Lambda))^m \right\}, \quad (21)$$

i.e. only  $k$  out of  $m$  sequences  $\{d_{\ell, \lambda}\}_{\lambda \in \Lambda}$  do not vanish. The space  $X_k$  is no longer a subspace since two different  $x$  might correspond to two different support sets  $\mathcal{S}$  and therefore its sum is not contained in  $X_k$ . The space  $X_k$  can be seen as a union of (shift invariant) subspaces.

To solve the recovery problem we propose a constrained optimization approach. Let therefore the linear sensing operator  $T$  be given by  $T : (\ell_2(\Lambda))^m \rightarrow (\ell_2(\Lambda))^p$  via  $Td = T(\{d_{1, \lambda}\}_{\lambda \in \Lambda}, \dots, \{d_{m, \lambda}\}_{\lambda \in \Lambda})^T = (\{(Ad_\lambda)^1\}_{\lambda \in \Lambda}, \dots, \{(Ad_\lambda)^p\}_{\lambda \in \Lambda})^T$ . For the purpose of identifying the support set  $\mathcal{S}$  we restrict the minimization of  $\|y^\delta - Td\|_{(\ell_2(\Lambda))^p}^2$  to the sub-domain

$$B(\Psi_{1,2}, R) = \{d \in (\ell_2(\Lambda))^m : \Psi_{1,2}(d) \leq R\},$$

where  $\Psi_{q,r}$  is a joint sparsity measure defined by  $\Psi_{q,r}(d) = \left( \sum_{\ell=1}^m \left( \sum_{\lambda \in \Lambda} |d_{\ell,\lambda}|^r \right)^{\frac{q}{r}} \right)^{\frac{1}{q}}$ . This measure forces the solution  $d$  for reasonably small chosen  $q$  (e.g.  $1 \leq q < 2$ ) to have non-vanishing rows  $\{d_{\ell,\lambda}\}_{\lambda \in \Lambda}$  only if  $\|\{d_{\ell,\lambda}\}_{\lambda \in \Lambda}\|_{\ell_r(\Lambda)}$  is large enough. Consequently, the optimization reads then as

$$\min_{d \in B(\Psi_{1,2}, R)} \|y^\delta - Td\|_{(\ell_2(\Lambda))^p}^2, \quad (22)$$

where the minimizing element in  $B(\Psi_{1,2}, R)$  can be approached by

$$d^{n+1} = \mathcal{P} \left( d^n + \frac{\gamma}{C} T^*(y^\delta - Td^n) \right), \quad (23)$$

where  $\gamma > 0$  is a step-length control (determined below) and  $\mathcal{P}$  is the  $\ell_2$ -projection on  $B(\Psi_{1,2}, R)$ , which can be realized by the sequence-valued generalized soft-shrinkage operator. To control the speed of convergence we introduce conditions on  $\gamma$ .

**Definition 2.** We say that the sequence  $\{\gamma^n\}_{n \in \mathbb{N}}$  satisfies Condition (B) with respect to the sequence  $\{d^n\}_{n \in \mathbb{N}}$  if there exists  $n_0$  such that:

$$\begin{aligned} \text{(B1)} \quad & \sup\{\gamma^n; n \in \mathbb{N}\} < \infty \quad \text{and} \quad \inf\{\gamma^n; n \in \mathbb{N}\} \geq 1 \\ \text{(B2)} \quad & \gamma^n \|Td^{n+1} - Td^n\|_{(\ell_2(\Lambda))^p}^2 \leq C \|d^{n+1} - d^n\|_{(\ell_2(\Lambda))^m}^2 \quad \forall n \geq n_0. \end{aligned}$$

**Proposition 4.** If for arbitrarily chosen  $d^0$  assume  $d^{n+1}$  is given by

$$d^{n+1} = \mathcal{P} \left( d^n + \frac{\gamma^n}{C} T^*(y^\delta - Td^n) \right), \quad (24)$$

with  $\gamma^n$  satisfying Condition (B) with respect to  $\{d^n\}_{n \in \mathbb{N}}$ , the sequence of residuals  $\|y^\delta - Td^n\|_{(\ell_2(\Lambda))^p}^2$  is monotonically decreasing and  $\{d^n\}_{n \in \mathbb{N}}$  converges in norm towards  $d^*$ , where  $d^*$  fulfills the necessary condition for a minimum of (22).

### 3.3 Compressive sensing and recovery for ill-posed problems

The objective in the sensing scenario for ill-posed problems is again to recover  $x$ , but now we only have access to  $Kx$  and  $K$  is supposed to be a linear (possibly ill-posed) and bounded operator between Hilbert spaces  $X$  and  $Y$ .

The data  $y$  are obtained by sensing  $Kx$  through  $F_s : Y \rightarrow (\ell_2(\Lambda))^p$ , i.e.  $y = F_s Kx = F_s K F_a^* d$ . Similarly to Lemma 1, we have the following result.

**Lemma 2.** Assume for all  $\lambda \in \Lambda$  the sampling functions  $s_{1,\lambda}, \dots, s_{p,\lambda}$  are chosen as in (14). Then  $y = A F_{K^* \nu} F_a^* d = A F_\nu F_{K_a}^* d$ .

An ideal choice to guarantee recovery within the compressive sampling framework would be to ensure  $F_{K^*v}F_a^* = F_vF_{Ka}^* = Id$ . For normalized systems  $\Phi_a$  and  $\Phi_v$  and ill-posed operators  $K$  this is impossible to achieve. The simplest case is that we have systems  $\Phi_a$  and  $\Phi_v$  at our disposal that diagonalize  $K$ , i.e.  $\langle Ka_{\ell,\lambda}, v_{\ell',\lambda'} \rangle = \kappa_{\ell,\lambda} \delta_{\lambda'\lambda} \delta_{\ell'\ell}$ . One prominent example is the so-called wavelet-vaguelette decomposition with respect to  $K$ . If  $\Phi_a$  and  $\Phi_v$  diagonalize  $K$ , then the structure of the sensing operator is  $TD : (\ell_2(\Lambda))^m \rightarrow (\ell_2(\Lambda))^p$ , where  $(TD)(\{d_{1,\lambda}\}_{\lambda \in \Lambda}, \dots, \{d_{m,\lambda}\}_{\lambda \in \Lambda}) = (\{(AD_\lambda d_\lambda)^1\}_{\lambda \in \Lambda}, \dots, \{(AD_\lambda d_\lambda)^p\}_{\lambda \in \Lambda})$ , and  $D$  is defined by  $\lambda$ -dependant blocks  $D_\lambda$  of size  $m \times m$ ,  $D_\lambda = \text{diag}(\kappa_{1,\lambda}, \kappa_{2,\lambda}, \dots, \kappa_{m,\lambda})$ .

Let us first consider the sensing problems for each individual label  $\lambda$  (which are  $m$ -dimensional recovery problems),

$$y_\lambda^\delta = AD_\lambda d_\lambda + z_\lambda \quad \text{with} \quad \|z_\lambda\| \leq \delta. \quad (25)$$

Since  $K$  is ill-posed, the sensing matrix  $AD_\lambda$  obeys no longer the restricted isometry property. Therefore, we propose to minimize the stabilized constrained optimization problem

$$\min_{d_\lambda \in B(\ell_1, R)} \|y_\lambda^\delta - AD_\lambda d_\lambda\|_{\ell_2}^2 + \alpha \|d_\lambda\|_{\ell_2}^2, \quad (26)$$

where  $B(\ell_1, R) = \{d_\lambda \in \ell_2 : \|d_\lambda\|_{\ell_1} \leq R\}$ . Let us define  $L^2 := D_\lambda A^* AD_\lambda + \alpha I$ , if  $A$  fulfills the restricted isometry property (16), then the operator  $L$  obeys a restricted isometry condition of the following form,

$$(\kappa_{\min}^2(1 - \delta_k) + \alpha) \|d_\lambda\|_{\ell_2}^2 \leq \|Ld_\lambda\|_{\ell_2}^2 \leq (\kappa_{\max}^2(1 + \delta_k) + \alpha) \|d_\lambda\|_{\ell_2}^2, \quad (27)$$

for all  $k$ -sparse vectors  $d_\lambda$  and where  $\kappa_{\max}$  denotes the largest and  $\kappa_{\min}$  the smallest eigenvalue of  $D_\lambda$ .

**Theorem 9 (finite dimensions).** *Assume  $R$  is such that  $d_\lambda \notin B(\ell_1, R)$  and that*

$$0 \leq \delta_{2k} < \frac{(1 + \sqrt{2})\kappa_{\min}^2 - \kappa_{\max}^2 + \sqrt{2}\alpha}{(1 + \sqrt{2})\kappa_{\min}^2 + \kappa_{\max}^2}. \quad (28)$$

*Then the minimizer  $d_\lambda^*$  of (26) satisfies*

$$\|d_\lambda^* - d_\lambda\|_{\ell_2} \leq C_0 k^{-1/2} \|d_\lambda^k - d_\lambda\|_{\ell_1} + C_1 \|L(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + C_2 \delta + C_3 \sqrt{\alpha} R, \quad (29)$$

*where  $d_\lambda^\dagger$  is the  $B(\ell_1, R)$ -best approximate solution,  $d_\lambda^k$  the best  $k$ -term approximation, and where the constants  $C_0$ ,  $C_1$ ,  $C_2$ , and  $C_3$  are given explicitly.*

As (28) serves as a condition for  $\delta_{2k}$  and  $\alpha$  at the same time, it turns out that the choice of  $\alpha$  influences the choice of a suitable sensing matrix  $A$  and vice versa.

Let us now investigate the full infinite dimensional measurement model,

$$y^\delta = (TD)d + z \quad \text{with} \quad \|z\|_{(\ell_2(\Lambda))^m} \leq \delta.$$

We propose to solve the following optimization problem,

$$\min_{d \in B(\Psi_{1,2}, R)} \|y^\delta - (TD)d\|_{(\ell_2(\Lambda))^p}^2 + \alpha \|d\|_{(\ell_2(\Lambda))^m}^2. \quad (30)$$

For the minimizing element the following error estimate hold true.

**Theorem 10 (infinite dimensions).** *Assume  $R$  is such that  $d \notin B(\Psi_{1,2}, R)$  and  $\delta_{2k}$  is as in Theorem 9. Then the minimizer  $d^*$  of (30) satisfies*

$$\|d^* - d\|_{(\ell_2(\Lambda))^m} \leq C_0 k^{-1/2} \Psi_{1,2}(d^k - d) + C_1 \|L(d^\dagger - d)\|_{(\ell_2(\Lambda))^m} + C_2 \delta + C_3 \sqrt{\alpha} R.$$

The minimizing elements can be iteratively approximated by

$$d_\lambda^{n+1} = \mathcal{P} \left( D_\lambda A^* (y_\lambda^\delta - AD_\lambda d_\lambda^n) \frac{\gamma^n}{C} + \left( 1 - \frac{\alpha \gamma^n}{C} \right) d_\lambda^n \right)$$

for problem (26) and for the full infinite dimensional case by

$$d^{n+1} = \mathcal{P} \left( D^* T^* (y^\delta - TDd^n) \frac{\gamma^n}{C} + \left( 1 - \frac{\alpha \gamma^n}{C} \right) d^n \right).$$

The norm convergence is ensured by Proposition 4.

## References

1. Adcock, B., Hansen, A.C.: Generalized sampling and infinite dimensional compressed sensing. Technical report NA2011/02, DAMTP, University of Cambridge (2012)
2. Adcock, B., Hansen, A.C., Herrholz, E., Teschke, G.: Generalized sampling, infinite-dimensional compressed sensing, and semi-random sampling for asymptotically incoherent dictionaries. submitted (2012)
3. Adcock, B., Hansen, A.C., Herrholz, E., Teschke, G.: Generalized sampling: extension to frames and inverse and ill-posed problems. *Inverse Problems* **29**(1), 015,008 (2013)
4. Adcock, B., Hansen, A.C., Roman, B., Teschke, G.: Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. to appear in *Adv. in Imag. and Electr. Phys.*, (arXiv:1310.1141v1) (2013)
5. Bonesky, T., Bredies, K., Lorenz, D.A., Maass, P.: A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Problems* **23**, 2041–2058 (2007)
6. Bredies, K., Lorenz, D., Maass, P.: A generalized conditional gradient method and its connection to an iterative shrinkage method. *Computational Optimization and Application* **42**(2), 173–193 (2009)
7. Bredies, K., Lorenz, D.A.: Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications* **14**(5–6), 813–837 (2008)
8. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory* **52**(2), 489–509 (2006)
9. Candès, E., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* **59**(8), 1207–1223 (2006)
10. Candès, E., Tao, T.: Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory* **52**(12), 5406–5425 (2006)

11. Candès, E.J.: The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences* **346**(9-10), 589–592 (2008)
12. Candès, E.J., Tao, T.: Decoding by linear programming. *IEEE Transaction on Information Theory* **51**(12), 4203–4215 (2005)
13. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Review* **43**(1), 129–159 (2001)
14. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **57**(11), 1413–1457 (2004)
15. Daubechies, I., Teschke, G., Vese, L.: Iteratively solving linear inverse problems with general convex constraints. *Inverse Problems and Imaging* **1**(1), 29–46 (2007)
16. Daubechies, I., Teschke, G., Vese, L.: On some iterative concepts for image restoration. *Advances in Imaging and Electron Physics* **150**, 1–51 (2008)
17. Denis, L., Lorenz, D.A., Dennis, T.: Greedy solution of ill-posed problems: error bounds and exact inversion. *Inverse Problems* **25**(11), 115,017 (24pp) (2009)
18. Donoho, D.: Compressed sensing. *IEEE Transactions on Information Theory* **52**(4), 1289–1306 (2006)
19. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell^1$  minimization. *Proc. Natl. Acad. Sci.* **100**, 2197–2202 (2003)
20. Donoho, D.L., Hou, X.: Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory* **47**, 2845–2862 (2001)
21. Dossal, C., Mallat, S.: Sparse spike deconvolution with minimum scale. In: *Proceedings of SPARS05* (2005)
22. Eldar, Y.C.: Compressed Sensing of Analog Signals in Shift-Invariant Spaces. *IEEE Trans. on Signal Processing* **57**(8), 2986–2997 (2009)
23. Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems*. Kluwer, Dordrecht (1996)
24. Feuer, A., Nemirovski, A.: On sparse representations in unions of bases. *IEEE Trans. Inform. Theory* **49**, 1579–1581 (2003)
25. Fornasier, M.: Domain decomposition methods for linear inverse problems with sparsity constraints. *Inverse Problems* **23**, 2505–2526 (2007)
26. Fornasier, M., Rauhut, H.: Recovery algorithms for vector valued data with joint sparsity constraint. *SIAM J. Numer. Anal.* **46**(2), 577–613 (2008)
27. Grasmair, M., Haltmeier, M., Scherzer, O.: Sparse regularization with  $\ell^q$  penalty term. *Inverse Problems* **24**(5), 055,020 (13pp) (2008)
28. Gribonval, R., Nielsen, M.: Beyond sparsity: Recovering structured representations by  $\ell^1$  minimization and greedy algorithms. *Advances in Computational Mathematics* **28**(1), 23–41 (2008)
29. Groetsch, C.W.: *The Theory of Tikhonov regularization for Fredholm Equations of the First Kind*. Pitman, Boston (1984)
30. Groetsch, C.W.: *Inverse Problems in the Mathematical Sciences*. Vieweg, Braunschweig (1993)
31. Herrholz, E., Teschke, G.: Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. *Inverse Problems* **26**, 125,012 (2010)
32. Klann, E., Kuhn, M., Lorenz, D.A., Maass, P., Thiele, H.: Shrinkage versus deconvolution. *Inverse Problems* **23**, 2231–2248 (2007)
33. Laska, J., S.Kirolos, Duarte, M., Ragheb, T., Baraniuk, R., Massoud, Y.: Theory and implementation of an analog-to-information converter using random demodulation. In: *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, New Orleans (2007)
34. Lorenz, D.A.: Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *Journal of Inverse and Ill-Posed Problems* **16**(5), 463–478 (2008)
35. Lorenz, D.A., Schiffler, S., Tredde, D.: Beyond convergence rates: Exact inversion with Tikhonov regularization with sparsity constraints. *Inverse Problems* **27**, 085,009 (2011)
36. Mallat, S., Davis, G., Zhang, Z.: Adaptive time-frequency decompositions. *SPIE Journal of Optical Engineering* **33**(7), 2183–2191 (1994)

37. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* **41**(12), 3397–3415 (1993)
38. Mishali, M., Eldar, Y.C.: Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. on Signal Processing* **57**(3), 993–1009 (2009)
39. Morozov, V.A.: *Methods for Solving Incorrectly Posed Problems*. Springer, New York (1984)
40. Pati, Y., Rezaiifar, R., Krishnaprasad, P.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 40–44 (1993)
41. Ramlau, R., Teschke, G.: A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numerische Mathematik* **104**(2), 177 – 203 (2006)
42. Rauhut, H., Schass, K., Vandergheynst, P.: Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory* **54**(5), 2210–2219 (2008)
43. Soulez, F., Denis, L., Fournier, C., Thiébaud, É., Goepfert, C.: Inverse problem approach for particle digital holography: accurate location based on local optimisation. *Journal of the Optical Society of America A* **24**(4), 1164–1171 (2007)
44. Teschke, G.: Multi-frame representations in linear inverse problems with mixed multi-constraints. *Appl. Computat. Harmon. Anal.* **22**, 43 – 60 (2007)
45. Teschke, G., Borries, C.: Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints. *Inverse Problems* **26**, 025,007 (2010)
46. Tropp, J.A.: Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory* **50**(10), 2231–2242 (2004)