

# A Note on Sparse Reconstruction Methods

**Martin Burger**

Westfälische Wilhelms-Universität Münster, Institut für Numerische und Angewandte  
Mathematik, Einsteinstr. 62, D 48149 Münster, Germany.

E-mail: [martin.burger@uni-muenster.de](mailto:martin.burger@uni-muenster.de)

**Abstract.** In this paper we discuss some aspects of sparse reconstruction techniques for inverse problems, which recently became popular due to several superior properties compared to linear reconstructions. We briefly review the standard sparse reconstructions based on  $\ell^1$ -minimization of coefficients with respect to an orthonormal basis, and also some recently proposed improvements based on Bregman iterations and inverse scale space techniques. For the latter we provide uniqueness results not available before for inverse problems with sparsity constraints.

In order to gain further understanding of sparse reconstruction techniques we provide a detailed analysis in the singular basis for the operator describing the inverse problem. This allows to compute analytic expressions for the reconstructions and highlight certain features. We also show that a very classical linear reconstruction technique, the truncated singular value decomposition is indeed equivalent to a sparse reconstruction technique with data-dependent weights.

Finally we touch the question whether it pays off to use sparse reconstruction schemes directly for the full inverse problem or if simple two-step schemes, consisting of a linear inversion and subsequent shrinkage, can potentially yield results of similar quality.

## 1. Introduction

Sparse reconstruction schemes, in particular regularization based on singular energies, have become very popular for the solution of linear and nonlinear inverse problems over the last years. The word sparsity usually refers to a low number of nonzero coefficients of the reconstructions in a certain basis, but is also used (with similar methods) with respect to other structures, e.g. sparse gradients in total variation methods. We refer to [23, 12, 8, 6, 10, 11] for classical results and recent developments.

In this paper we highlight some aspects of sparse reconstruction methods for linear inverse problems, described by an operator  $A : X \rightarrow Y$  between Hilbert spaces  $X$  and  $Y$ . We seek sparse solutions in an orthonormal basis  $\{\phi_j\}_{j \in \mathbb{N}}$ . The standard approach is the weighted  $\ell^1$ -minimization

$$\frac{1}{2} \left\| \sum_j c_j A \phi_j - y \right\|^2 + \alpha \sum_j w_j |c_j| \rightarrow \min_{(c_j) \in \ell_{\mathbf{w}}^1(\mathbb{N}) \cap \ell^2(\mathbb{N})}, \quad (1.1)$$

with positive weights  $\mathbf{w} = (w_i)$ . Here  $\ell_{\mathbf{w}}^1(\mathbb{N})$  denotes the space of coefficients  $c_j$  such that

$$\sum_j w_j |c_j| < \infty.$$

Under appropriate conditions on the weights (in particular if they are bounded away from zero), the existence of a solution can be guaranteed, and uniqueness holds at least if  $A$  has trivial nullspace. The optimality condition for this convex optimization problem takes the form

$$\langle A\phi_i, \sum_j c_j A\phi_j - y \rangle + \alpha w_i s_i = 0, \quad s_i \in \text{sign}(c_i), \quad i \in \mathbb{N}, \quad (1.2)$$

where we define the multi-valued sign as

$$\text{sign}(c) = \begin{cases} 1 & \text{if } c > 0 \\ [-1, 1] & \text{if } c = 0 \\ -1 & \text{if } c < 0. \end{cases} \quad (1.3)$$

The basic assumptions we use in this paper is that  $A$  is a compact operator acting from a separable Hilbert space  $X$  (with orthonormal basis  $\{\phi_j\}$ ) to a Hilbert space  $Y$ . We assume for simplicity in the remainder of the paper also that  $A$  has trivial nullspace, including a nullspace of the operator can be performed in a straight-forward way for most results with only additional technical effort. Under these conditions it is easy to prove that (1.1) has a unique minimizer, noticing that the first term is strictly convex and weakly continuous on  $\ell^2(\mathbb{N})$  (due to the compactness of  $A$ ) and the second term is convex and weakly lower semicontinuous. In order to simplify the analysis we also introduce the operator  $B : \ell^2(\mathbb{N}) \rightarrow Y$ ,

$$(c_j) \mapsto \sum_j c_j A\phi_j.$$

The  $\ell^1$ -minimization can then be rewritten as

$$\frac{1}{2} \|B(c_j) - y\|^2 + \alpha \sum_j w_j |c_j| \mapsto \min_{(c_j) \in \ell_w^1(\mathbb{N}) \cap \ell^2(\mathbb{N})}, \quad (1.4)$$

with the optimality condition  $B^*B(c_j) + \alpha(w_j s_j) = 0$ ,  $B^* : Y \rightarrow \ell^2(\mathbb{N})$  denoting the adjoint operator of  $B$ . Moreover we will assume that  $w_j$  are positive weights and that there exists a constant  $\underline{w} > 0$  such that  $w_j \geq \underline{w}$  for all  $j \in \mathbb{N}$ . Hence  $\sum w_j |c_j|$  is really a norm on  $\ell^1(\mathbb{N})$ .

Recently novel multiscale reconstruction schemes based on Bregman iterations and inverse scale space methods have been introduced, which can improve some shortcomings of the variational schemes. We shall in particular focus on the inverse scale space method, which reads

$$w_i \frac{ds_i}{dt}(t) = -\langle A\phi_i, \sum_j c_j A\phi_j - y \rangle, \quad s_i(t) \in \text{sign}(c_i(t)) \quad (1.5)$$

in the above setup. We will discuss its derivation and basic properties in Section 2.

In this paper we try to contribute to the understanding of the sparse reconstruction schemes (1.1) and (1.5). After preliminary properties being reviewed in Section 2, we detail the error estimation in Bregman distance for the sparse reconstructions, which are obtained by specializing and interpreting general results obtained previously (cf. [4, 5]). In Section 4 we further investigate the special case of sparse reconstructions in the singular basis for the operator  $A$ , which allows to obtain explicit formulas and further insight. Moreover, we show that for some particular data-dependent weights in the  $\ell^1$ -regularization the classical truncated singular value decomposition is obtained as a special case. In Section 5 we shall use the results for the singular value case to compare direct sparse reconstructions with a two-step procedure, where first a linear reconstruction is computed and subsequent shrinkage is performed.

## 2. Bregman Iterations and Inverse Scale Space Methods

In the following we provide further details on the construction of sparse Bregman iterations and sparse inverse scale space methods. A major ingredient for the approach is the iterative approach is the *Bregman distance*, defined by

$$D_J^p(v, u) = J(v) - J(u) - \langle p, v - u \rangle,$$

for a convex functional  $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$  and a subgradient

$$p \in \partial J(u) = \{ q \in X^* \mid \langle q, v - u \rangle \leq J(v) - J(u) \}.$$

In the case of  $J$  being the norm in  $\ell_{\mathbf{w}}^1(\mathbb{N})$ -penalization, subgradients are of the form  $p = (w_j s_j)_{j \in \mathbb{N}}$  with  $s_j \in \text{sign}(c_j)$ . Hence, the Bregman distance for the coefficients  $c_j$  related to sparsity constraints (which we denote by  $D_S^p$  in the following) becomes

$$D_S^p((\gamma_j), (c_j)) = \sum_{j=1}^{\infty} w_j (|\gamma_j| - s_j \gamma_j), \quad s_j \in \text{sign}(c_j). \quad (2.1)$$

The sparsity of the reconstruction can be seen from the optimality condition (1.2), respectively its equivalent statement in terms of the operator  $B$ , given by

$$B^*(B(c_j) - y) + \alpha(w_j s_j) = 0. \quad (2.2)$$

We have  $(w_j s_j)$  in the range of  $B^* \subset \ell^2(\mathbb{N})$ , thus

$$\underline{w} \sum_j s_j^2 \leq \sum_j w_j^2 s_j^2 < \infty.$$

This is only possible if  $s_j^2 < 1$  except for a final number of  $j$ . Hence, only a final number of  $c_j$  are nonzero.

We can now generalize the variational approach (1.1) to iterated regularization approaches, generating a sequence  $(c_j^k)$ , with appropriate choice of the iteration number  $k$  in order to achieve a regularizing effect. In the *Bregman iteration*, the regularization term (the convex functional  $J$ ) is replaced by the Bregman distance to the last iterate. In the case of sparsity constraints this means that the coefficients  $(c_j^{k+1})$  are obtained via minimizing the convex functional

$$\frac{1}{2} \left\| \sum_j c_j A \phi_j - y \right\|^2 + \alpha \sum_j w_j (|c_j| - c_j s_j^k) \rightarrow \min_{c_j \in \ell_{\mathbf{w}}^1(\mathbb{N}) \cap \ell^2(\mathbb{N})}, \quad (2.3)$$

for  $s_j^k \in \text{sign}(c_j^k)$ . The optimality condition for the Bregman iteration can be interpreted as an iteration formula for the subgradients, one observes

$$\langle A \phi_i, \sum_j c_j^{k+1} A \phi_j - y \rangle + \alpha w_i (s_i^{k+1} - s_i^k) = 0, \quad s_i^{k/k+1} \in \text{sign}(c_i^{k/k+1}), \quad i \in \mathbb{N}. \quad (2.4)$$

The standard starting value for the Bregman iteration is identical zero for the coefficients as well as for the  $s_i$ , which clearly guarantees to start with a subgradient.

It is particularly interesting to consider the overdamped case  $\alpha \rightarrow \infty$  with the interpretation  $s_i^k = s_i(\frac{k}{\alpha})$ , which leads a flow in pseudo time  $t \sim \frac{k}{\alpha}$ , the so-called inverse scale space method (cf. [24, 3]). One identifies the formal limit as (1.5), since

$$\alpha(s_i^{k+1} - s_i^k) \rightarrow \frac{ds_i}{dt}.$$

A detailed analysis of this limit and the existence and uniqueness of the flow is indeed a challenging problem, which has not yet been carried out in full generality (see the analysis in [3] and in [2] for the total variation case). The existence can be obtained by standard a-priori estimates and weak convergence techniques, whereas uniqueness is a non-trivial problem - in particular since the regularization term is not strictly convex. We therefore give a detailed uniqueness proof in the following:

**Theorem 2.1.** *Let  $A$  have trivial nullspace and  $w_i > 0$  for all  $i$ . Then there exists a unique solution  $(c_j(t))_{j \in \mathbb{N}} \in \ell^1_{\mathbf{w}}(\mathbb{N})$  for the sparse inverse scale space method (1.5) one solution.*

*Proof.* Let  $(c_i, s_i)_{i \in \mathbb{N}}$  and  $(\tilde{c}_i, \tilde{s}_i)_{i \in \mathbb{N}}$  be two solutions of (1.5), then their difference  $(\gamma_i, \sigma_i) = (c_i - \tilde{c}_i, s_i - \tilde{s}_i)$  satisfies

$$w_i \frac{d\sigma_i}{dt} = - \sum_j \gamma_j \langle A\phi_i, A\phi_j \rangle$$

with zero initial values of  $\sigma_i$  and  $c_i$ . Due to the trivial nullspace of  $A$  it is easy to see that the operator  $B$  as defined above has trivial nullspace and hence,  $B^*B$  is symmetric positive definite ( $\langle B^*Bu, u \rangle > 0$  for all  $u \in \ell^2(\mathbb{N}) \setminus \{0\}$ ). The above relation for differences shows that  $(w_i \frac{d\sigma_i}{dt})$  lies in the range of  $B^*B$ . Hence we can use the right-inverse of  $B$  (well-defined on its range) and see that

$$(B^*B)^{-1}(w_i \frac{d\sigma_i}{dt}) = -(\gamma_i),$$

and hence

$$\frac{d}{dt} \frac{1}{2} \|(B^*)^{-1}(w_i \sigma_i)\|_{\ell^2}^2 = \langle (w_i \sigma_i), B^{-1}(B^*)^{-1}(w_i \frac{d\sigma_i}{dt}) \rangle = -\langle w_i \sigma_i, \gamma_i \rangle = - \sum_i w_i (s_i - \tilde{s}_i)(c_i - \tilde{c}_i) \leq 0.$$

Hence  $(B^*)^{-1}(w_i \sigma_i) = 0$  and due to the properties of  $B$  and the positivity of the weights we conclude  $\sigma_i \equiv 0$ . Inserting  $\frac{d\sigma_i}{dt=0}$  in the flow equation implies

$$\sum_j \gamma_j \langle A\phi_i, A\phi_j \rangle = 0$$

and consequently

$$\| \sum_j \gamma_j A\phi_j \|_Y^2 = \langle \sum_i \gamma_i A\phi_i, \sum_j \gamma_j A\phi_j \rangle = 0$$

for all  $t$ . Due to the trivial nullspace of  $A$  and the fact that  $(\phi_j)$  is a basis of  $X$  we conclude  $\gamma_i(t) = 0$  for all  $i \in \mathbb{N}$  and all  $t \geq 0$ . Hence, the solution is unique.  $\square$

### 3. Error Estimation for Sparse Reconstructions

In this section we rephrase some previously obtained general results for error estimation of regularization methods with general convex regularization functionals (cf. [4, 5, 21]) in the case of sparsity constraints and we interpret the used error measures and conditions in this case.

The fundamental condition used for error estimation is the so-called *source condition*, which assumes that there exists a solution  $u$  for exact data such that there is a subgradient  $p \in \partial J(u)$ , which also lies in the range of the adjoint operator (of the one describing the inverse problem). In the context considered here, the source condition means

$$\exists z \in Y : (w_j \hat{s}_j) = B^*z, \quad \hat{s}_j \in \text{sign}(\hat{c}_j). \quad (3.1)$$

Note that by the same reasoning as for the optimality condition for  $\ell^1$ -minimization, the source condition implies that  $(\hat{c}_j)$  is sparse. We shall refer to  $\hat{y} = B(\hat{c}_j)$  as *exact data*.

The error estimation is carried out in the Bregman distance as defined above, so the results will mainly characterize the quality in the number of components and their sign. However, a recent result by Lorenz [18] shows that under reasonable condition on the operator  $A$  respectively  $B$ , estimates in the  $\ell^2$ -norm can be deduced in automatic way from the Bregman distance estimates. Error estimates for the  $\ell^1$ -minimization (1.1) (cf. [4] for general results on variational methods) can be obtained very easily by subtracting  $(w_j \hat{s}_j)$  in the optimality condition. Taking the inner product with  $(c_j - \hat{c}_j)$  this implies together with the source condition

$$\langle B(c_j - \hat{c}_j), B(c_j) - y \rangle + \alpha \sum_j w_j (s_j - \hat{s}_j) (c_j - \hat{c}_j) = -\alpha \sum_j w_j \hat{s}_j (c_j - \hat{c}_j) = -\alpha \langle z, B(c_j - \hat{c}_j) \rangle.$$

In the case of exact data ( $y = \hat{y} = B(\hat{c}_j)$ ) an error estimate can now be concluded applying Young's inequality on the right-hand side, which implies

$$\frac{1}{2\alpha} \|B(c - \hat{c}_j)\|_Y^2 + \sum_j w_j (s_j - \hat{s}_j) (c_j - \hat{c}_j) \leq \frac{\alpha}{2} \|z\|_Y^2 \quad (3.2)$$

A straight-forward calculation shows that

$$\sum_j w_j (s_j - \hat{s}_j) (c_j - \hat{c}_j) = D_S^p((\hat{s}_j), (s_j)) + D_S^{\hat{p}}((s_j), (\hat{s}_j)), \quad (3.3)$$

i.e. (3.2) implies a an estimate for both possible Bregman distances, being of order  $\alpha$ . Moreover, the difference in the weaker norm of the output space  $Y$ ,  $\|B(c - \hat{c}_j)\|_Y$ , is of order  $\alpha$ . Note that the above estimate holds for any subgradient  $(\hat{s}_j)$  satisfying the source condition, so that more specific results can be obtained by choosing appropriate subgradients. In particular it is often possible to choose  $\hat{s}_j = 0$  if  $\hat{c}_j = 0$  so that the Bregman distance will majorize the weighted  $\ell^1$ -norm of  $(c_j)$  on the zero components of  $\hat{c}_j$ . This means that smallness of the Bregman distance from the above estimate implies that the reconstructed coefficients  $c_j$  are small in absolute value if  $\hat{c}_j = 0$ .

In the case of noisy data, i.e. if  $y$  is a perturbation of exact data  $\hat{y} = B(\hat{c}_j)$ , bounded in the norm by some noise level  $\delta$ , i.e.

$$\|y - \hat{y}\| \leq \delta.$$

The estimate can then easily be modified to (cf. [5])

$$\frac{1}{2\alpha} \|B(c_j) - \hat{y}\|_Y^2 + \sum_j w_j (s_j - \hat{s}_j) (c_j - \hat{c}_j) \leq \frac{\alpha}{2} \|z\|_Y^2 + \frac{\delta^2}{\alpha}. \quad (3.4)$$

An analogous estimate can be made for inverse scale space methods. The major difference is that a one-sided Bregman distance has to be used, with the special subgradient  $s_j \in \text{sign}(c_j)$  constructed from the inverse scale space method, and in particular the estimate is not multivalued. The Bregman distance estimate for exact data (cf. [5]) then becomes

$$\sum_j w_j \hat{c}_j (\hat{s}_j - s_j) \leq \frac{\|z\|_Y^2}{t}, \quad (3.5)$$

and the noisy case can be handled in a similar way with an additional term depending on  $t\delta^2$  on the right-hand side. The interpretation of the estimate for the inverse scale space is more difficult than for the variational method, in particular it does not provide information on the behaviour for  $\hat{c}_j = 0$ . However, one expects an analogous behaviour of the inverse scale space method and the variational method in particular for those components, which will become more clear in the next section when sparse reconstructions in the orthogonal basis are discussed.

#### 4. Sparsity in the Singular Basis

In the following we consider the simplifying case of choosing the singular vectors of the linear operator  $A$  as a basis for sparse reconstruction. More precisely, let  $(\phi_k)_{k \in \mathbb{N}}$  and  $(\psi_k)_{k \in \mathbb{N}}$  be orthonormal bases of  $X$  and  $Y$ , respectively, such that

$$A\phi_k = \sigma_k\psi_k, \quad A\psi_k = \sigma_k\phi_k \quad (4.1)$$

for the sequence  $(\sigma_k)_{k \in \mathbb{N}} \subset \mathbb{R}^+$  of singular values (cf. [] for the well-definedness of this singular decomposition for compact operators).

The sparse reconstruction considerably simplifies in the singular basis, due to the separability of the functionals involved in (1.1). Due to the pairwise orthogonality of the  $A\phi_k$  the weighted  $\ell^1$ -penalized problem can be rewritten as

$$\frac{1}{2} \sum_j (\sigma_j c_j - y_j)^2 + \alpha \sum_j w_j |c_j| \rightarrow \min_{(c_j) \in \ell_{\mathbf{w}}^1(\mathbb{N}) \cap \ell^2(\mathbb{N})}. \quad (4.2)$$

with  $y_j := \langle y, \psi_j \rangle$ . Hence, the minimizer can be computed by minimizing the functional  $E_j : \mathbb{R} \rightarrow \mathbb{R}^+$ .

$$E_j(c) = \frac{1}{2}(\sigma_j c - y_j)^2 + \alpha w_j |c| \quad (4.3)$$

to obtain  $c_j$  and a final check that  $(c_j) \in \ell_{\mathbf{w}}^1(\mathbb{N}) \cap \ell^2(\mathbb{N})$  holds, which becomes a standard shrinkage problem (see below for details).

With analogous computation, the inverse scale space method becomes

$$w_i \partial_t s_i(t) = -\sigma_i(\sigma_i c_i - y_i), \quad s_i(t) \in \text{sign}(c_i(t)), \quad c_i(0) = s_i(0) = 0. \quad (4.4)$$

One immediately observes that the evolution equations for  $c_i$  and  $s_i$  decouple with respect to the index  $i$ , so again one only needs to solve scalar problems for each component, which will lead to a different shrinkage procedure as we shall show below.

**Proposition 4.1.** *The unique minimizer of (4.2) is given by*

$$c_j^\alpha = \begin{cases} \text{sign}(y_j) \frac{\sigma_j |f_j| - \alpha w_j}{\sigma_j^2} & \text{if } \sigma_j |y_j| \geq \alpha w_j \\ 0 & \text{else} \end{cases} \quad (4.5)$$

The unique solution of (4.4) is given by

$$c_j(t) = \begin{cases} \frac{y_j}{\sigma_j} & \text{if } \sigma_j |y_j| \geq \frac{w_j}{t} \\ 0 & \text{else} \end{cases}, \quad s_j(t) = \begin{cases} \frac{\sigma_j y_j t}{w_j} & \text{if } \sigma_j |y_j| < \frac{w_j}{t} \\ \text{sign}(y_j) & \text{else} \end{cases} \quad (4.6)$$

*Proof.* As noticed above, the minimizers  $c_j^\alpha$  of (4.2) can be computed by minimizing the one-dimensional functionals  $E_j$ , with the optimality condition

$$-\frac{\sigma_j}{w_j}(\sigma_j c_j^\alpha - y_j) \in \text{sign}(c_j^\alpha).$$

The above formula for the optimal  $c_j^\alpha$  can be easily deduced by a case study with respect to the sign of  $c_j^\alpha$ .

Now consider (4.4) for  $t \leq \frac{\sigma_j |y_j|}{w_j}$ . Then  $s_j(t)$  given by the above formula and  $c_j(t)$  satisfy

$$w_j \partial_t c_j(t) = \sigma_j y_j = -\sigma_j^2 c_j(t) + \sigma_j y_j.$$

Moreover,  $|s_j(t)| \leq 1$  and hence,  $s_j(t) \in \text{sign}(c_j(t)) = \text{sign}(0)$ . For  $t > \frac{\sigma_j |y_j|}{w_j}$  we see  $s_j(t) \in \text{sign}(f_j) = \text{sign } c_j(t)$  and

$$w_j \partial_t c_j(t) = 0 = -\sigma_j^2 c_j(t) + \sigma_j y_j.$$

Hence, the evolution equation (4.4) is satisfied by  $c_j(t), s_j(t)$  in both cases.  $\square$

From Proposition 4.1 we see that the  $\ell^1$ -penalization can be seen as a version of soft shrinkage, while the inverse scale space method is a version of hard shrinkage for the inverse problem. In the case of  $A$  being the identity (all  $\sigma_j = 1$ ), this result is well-known (cf. [8, 26]).

We now examine the formula for  $c_j(t)$  solving (4.4) for a special choice of weights, which allows to obtain a well-known linear regularization method, the *truncated singular value decomposition*, as a special case. If we choose  $w_j = \frac{|y_j|}{\sigma_j^m}$ , then the truncation condition compares  $\sigma_j|y_j|$  and  $\frac{|y_j|}{t\sigma_j^m}$ , hence the solution of the sparse inverse scale space (4.4) becomes

$$c_j(t) = \begin{cases} \frac{y_j}{\sigma_j} & \text{if } \sigma_j^{m+1} \geq \frac{1}{t} \\ 0 & \text{else} \end{cases} \quad (4.7)$$

which is just truncated singular value decomposition if we take the  $(m+1)$ -th root in the truncation condition, i.e.:

**Corollary 4.2.** *The sparse inverse scale space method for the singular vectors  $\phi_i$  and the data-dependent weights  $w_i = \frac{|y_i|}{\sigma_i^m}$  is equivalent to the truncated singular value decomposition with truncation parameter  $\alpha = \frac{1}{t^{1/(1+m)}}$ .*

## 5. Direct Sparse Reconstruction vs. Two-Step Procedures

In the following we try to gain some further understanding of the comparison between direct sparse reconstructions and two-step procedures based on linear reconstruction followed by shrinkage. For this sake we shall use again the singular basis of the operator as in the previous section, which allows to gain insight into the properties of the schemes for different frequencies. We shall compare the direct reconstruction via the inverse scale space method (for simplicity with weights  $w_j = 1$ )

$$\frac{ds_j}{dt} = -(B^*(B(c_k) - y))_j, \quad s_j \in \text{sign}(c_j), \quad (5.1)$$

and the two step procedure

$$\frac{d\tilde{s}_j}{dt} = -\tilde{c}_j + d_j, \quad (d_j) = R_\alpha y, \quad (5.2)$$

where  $R_\alpha$  is a linear regularization with parameter  $\alpha > 0$  approximating the (generalized) inverse of  $B$ . A similar analysis with analogous results can be carried out for variational method, they just introduce an additional systematic error.

In the singular basis as above, a linear regularization can be written in the form (cf. [14])

$$(R_\alpha y)_j = r_\alpha(\sigma_j)y_j,$$

with a function  $r_\alpha(\sigma)$  converging (pointwise) to  $\frac{1}{\sigma}$  as  $\sigma \rightarrow 0$ . Then we can write the two-step approach in the condensed form

$$\frac{d\tilde{s}_j}{dt} = -\tilde{c}_j + r_\alpha(\sigma_j)y_j. \quad (5.3)$$

The solution of (5.3) can be computed explicitly as in the previous section

$$\tilde{c}_j(\tilde{t}) = \begin{cases} r_\alpha(\sigma_j)y_j & \text{if } r_\alpha(\sigma_j)|y_j| \geq \frac{1}{\tilde{t}} \\ 0 & \text{else} \end{cases}, \quad \tilde{s}_j(\tilde{t}) = \begin{cases} r_\alpha(\sigma_j)y_j\tilde{t} & \text{if } r_\alpha(\sigma_j)|y_j| < \frac{1}{\tilde{t}} \\ \text{sign}(y_j) & \text{else} \end{cases} \quad (5.4)$$

In an analogous way, the direct sparse reconstruction can be written as above

$$\frac{ds_j}{dt} = -\sigma_j^2 c_j + \sigma_j y_j,$$

with solution given by formula (4.6) for  $w_j = 1$ .

Using these explicit relations we can reasonably compare the behaviour of the two methods in different regimes:

- For  $\sigma_j$  large (compared to the regularization parameter) all standard linear regularization methods behave as  $r_\alpha(\sigma_j) \approx \sigma_j$ . Hence, the two-step scheme provides a reconstruction close to  $\frac{y_j}{\sigma_j}$  if

$$|y_j| \geq \frac{1}{t r_\alpha(\sigma_j)} \approx \frac{\sigma_j}{t}.$$

On the other hand, the direct sparse reconstruction yields  $\frac{y_j}{\sigma_j}$  if  $|y_j| \geq \frac{1}{\sigma_j t}$ . Hence, the latter will not introduce a significant error on the large frequencies ( $\frac{1}{\sigma_j t}$  will be rather small), while the two-step procedure can easily cut large frequency parts ( $\frac{\sigma_j}{t}$  can be of order one).

- For  $\sigma_j$  small (compared to the regularization parameter) linear regularizations yield a bounded approximation of the inverse, and hence  $r_\alpha(\sigma_j)$  behaves like  $\frac{1}{\alpha}$  (if the scaling of the regularization parameter is chosen appropriately). Hence, the reconstruction of the two-step scheme is  $\tilde{c}_j \approx \frac{y_j}{\alpha}$  if

$$|y_j| \geq \frac{1}{t r_\alpha(\sigma_j)} \approx \frac{\alpha}{t}.$$

The direct sparse reconstruction yields  $c_j = \frac{y_j}{\sigma_j}$  if  $|y_j| \geq \frac{\sigma_j}{t} \ll \frac{\alpha}{t}$ . Hence, the direct sparse reconstructions potentially reconstructs also smaller components, and in addition the two-step procedure usually introduces a systematic error from the linear reconstruction (division by  $\alpha \gg \sigma_j$ ). This property again favours the direct sparse reconstruction, in particular for small noise levels.

These arguments indicate that a two-step reconstruction will lead to a higher error even in the case of reconstruction in the singular basis, which should be the most favourable one for two-step methods. The linear reconstruction can always be interpreted as a reconstruction in the singular basis, and hence in the general case the second step will consist of a conversion to a different basis followed by shrinkage, which creates additional information loss.

## References

- [1] L.M.Bregman, *The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. Math. Phys. **7** (1967), 200-217.
- [2] M.Burger, K.Frick, S.Osher, O.Scherzer, *Inverse total variation flow*, SIAM Multiscale Modelling and Simulation **6** (2007), 366-395.
- [3] M.Burger, G.Gilboa, S.Osher, J.Xu, *Nonlinear inverse scale space methods*, Comm. Math. Sci. **4** (2006), 179-212.
- [4] M.Burger, S.Osher, *Convergence rates for convex variational regularization*, Inverse Problems **20** (2004), 1411-1421.
- [5] M.Burger, E.Resmerita, L.He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing (2007), to appear
- [6] E.J.Candes, T.Tao, *Decoding by linear programming*, IEEE Trans. Information Theory **51** (2005), 4203-4215.
- [7] A.Chambolle, *An algorithm for total variation regularization and denoising*, J. Math. Imaging Vision **20** (2004), 89-97.
- [8] A.Chambolle, R.DeVore, N.Y.Lee, B.Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Proc., **7** (1998), 319-335.

- [9] T.Chan, J.Shen, *Image Processing and Analysis* (SIAM, Philadelphia, 2005).
- [10] I.Daubechies, M.Defrise, C.DeMol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math. **57** (2004), 1413-1457.
- [11] D.Donoho, M.Elad, V.Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Information Theory **52** (2006), 618.
- [12] D.Donoho, I.Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika **81** (1994), 425-455.
- [13] I.Ekeland, R.Temam, *Convex analysis and variational problems*, Corrected Reprint Edition, SIAM, Philadelphia, 1999.
- [14] H.Engl, M.Hanke, A.Neubauer, *Regularization of Inverse Problems* (Kluwer, Dordrecht, 1996).
- [15] H.Y.Gao, A.G.Bruce, *WaveShrink with firm shrinkage*, Statist. Sinica 7 (1997), 855-874.
- [16] L.He, T.C.Chung, S.Osher, T.Fang, P.Speier, *MR Image Reconstruction by Using the Iterative Refinement Method and Nonlinear Inverse Scale Space Methods*, UCLA, CAM 06-35.
- [17] J.Lie, J.M. Nordbotten, *Inverse scale spaces for nonlinear regularization*, J. Math. Imaging Vision (2006), to appear.
- [18] D.Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, Preprint (University Bremen, 2008).
- [19] S.Osher, M.Burger, D.Goldfarb, J.Xu, W.Yin, *An iterative regularization method for total variation-based image restoration*, SIAM Multiscale Model. Simul. 4 (2005), 460-489.
- [20] S.Osher, J.Xu, *Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising*, IEEE Trans. Image Proc., to appear.
- [21] E.Resmerita, *Regularization of ill-posed problems in Banach spaces: convergence rates*, Inverse Problems 21 (2005) 1303-1314.
- [22] E.Resmerita, O.Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems 22 (2006) 801-814.
- [23] L.I.Rudin, S.J.Osher, E.Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259-268.
- [24] O.Scherzer, C.Groetsch, *Inverse scale space theory for inverse problems*, in: M.Kerckhove, ed., *Scale-space and Morphology in Computer Vision.Proceedings of the third International Conference Scale-space 2001* (Springer, Berlin, 2001), 317-325.
- [25] F.Schoepfer, A.K.Louis, T.Schuster, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces* Inverse Problems 22 (2006), 311-329.
- [26] J.Xu, S.Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising*, IEEE Trans. Image Proc. (2007), to appear.