On some iterative concepts for image restoration

Ingrid Daubechies
Princeton University, PACM
Washington Road
Princeton, NJ 08544-1000, USA

Gerd Teschke
Konrad–Zuse–Institute Berlin
Takustr. 7
D-14195 Berlin-Dahlem, Germany

Luminita Vese
Department of Mathematics, UCLA
405 Hilgard Avenue
Los Angeles, CA 90095-1555, USA

August 8, 2007

1   Introduction

In this paper, we discuss several iterative strategies for solving inverse problems in the context of signal and image processing. We are essentially focusing on problems where it is reasonable to assume that the solution has a sparse expansion with respect to a wavelet basis or frame. In each case, we consider a variational formulation of the problem, and construct an iteration scheme for which the iterates approximate the solution. To this end, we apply surrogate functionals; the corresponding strategy was shown to converge in norm and to regularize the problem, see (Daubechies et al., 2004). We discuss special cases and generalizations.

The surrogate functional method in its initial setup as described in (Daubechies et al., 2004) amounts to a combination of Landweber's method and a shrinkage operation, applied in each iteration step. The shrinkage is due to the presence of the ℓ_1-penalization term in the functional. Recent developments in the field of signal and image processing have shown the importance of sparse representations for various tasks in inverse problems (such as compression, denoising, deblurring, decomposition, texture analysis etc.); ℓ_1-constraints select for such sparsity. We limit ourselves here to illustrating a small number of concrete inverse problems, for which we show in detail the variational formulations and the resulting expressions for the iteration. For all cases we discuss convergence and

give detailed numerical illustrations. In addition to these case studies, we also present strategies for more general constraints.

We start with the concrete problem of simultaneously denoising, decomposing and deblurring a given image. The associated variational formulation of the problem contains terms that promote sparsity as well as smoothness. We show how to transform the problem such that the basic method of (Daubechies et al., 2004) applies. In a second example, we discuss a natural extension to vector-valued inverse problems. Potential applications include seismic or astrophysical data decomposition/reconstruction and color image reconstruction. The illustration presented here contains audio data coding. After these two case studies, we turn to more general formulations. We allow the constraint to be some other positive, homogeneous and convex functional than the ℓ_1-norm. In the linear case, and under fairly general assumptions on the constraint, we prove that weak convergence of the iterative scheme always holds. In certain cases, i.e. for special families of convex constraints, this weak convergence implies norm convergence. The presented technique covers a wide range of problems. Here we discuss in greater detail image restoration problems in which Besov– or BV–constraints are involved. We close this paper with sketching the design of hybrid wavelet–PDE image restoration schemes, i.e. with variational problems that contain wavelet as well as BV constraints.

2　Simultaneous decomposition, deblurring and denoising of images by means of wavelets

This section is devoted to wavelet–based treatments of variational problems arising in the field of image processing. In particular, we follow approaches presented in (Meyer, 2002; Vese and Osher, 2003, 2004; Osher et al., 2003) and discuss a special class of variational functionals that induce a decomposition of images into oscillating and cartoon components and possibly an appropriate 'noise' component. In the setting of (Vese and Osher, 2003) and (Osher et al., 2003), the cartoon component of an image is modelled by a BV function (bounded variation); the corresponding incorporation of BV penalty terms in the variational functional leads to PDE schemes that are numerically intensive. By replacing the BV penalty term by a $B_1^1(L_1)$ term (which amounts to a slightly stronger constraint on the minimizer), and writing the problem in a wavelet framework, we obtain elegant and numerically efficient schemes with results very similar to those obtained in (Osher et al., 2003), and superior to those from (Rudin et al., 1992). This approach allows us, moreover, to incorporate bounded linear blur operators into the problem so that the minimization leads to a simultaneous decomposition, deblurring and denoising.

2.1 Wavelet–based reformulation of the variational problem

As mentioned above, we focus on a special class of variational problems which induce a decomposition of images into 'texture' and 'cartoon' components; the cartoon part is, ideally, piecewise smooth with possibly abrupt edges and contours; the texture part on the other hand 'fills' in the smooth regions in the cartoon with, typically, oscillating features. Inspired from (Meyer, 2002), the authors of (Vese and Osher, 2003; Osher et al., 2003) propose to model the cartoon component by the space BV; this induces a penalty term that allows edges and contours in the reconstructed cartoon images, leading however to a numerically intensive PDE based scheme.

Our main goal is to provide a computationally thriftier algorithm by using a wavelet–based scheme that solves not the same but a very similar variational problem, in which the BV–constraint, which cannot easily be expressed in the wavelet domain, is replaced by a $B_1^1(L_1)$–term, i.e. a slightly stricter constraint (since $B_1^1(L_1) \subset BV$ in two dimensions). Moreover, we can easily incorporate the action of linear bounded blur operators; we also show convergence of the proposed scheme.

In order to give a brief description of the underlying variational problems, we recall the methods proposed in (Vese and Osher, 2003; Osher et al., 2003). They follow the idea of Y. Meyer (Meyer, 2002), proposed as an improvement on the total variation framework of (Rudin et al., 1992). In principle, the models can be understood as a decomposition of an image f into $f = u + v$, where u represents the cartoon part and v the texture part. In the (Vese and Osher, 2003, 2004) model, the decomposition is induced by solving

$$\inf_{u,g_1,g_2} G_p(u, g_1, g_2) \ , \quad \text{where} \tag{2.1}$$

$$G_p(u, g_1, g_2) = \int_\Omega |\nabla u| + \lambda \|f - (u + \operatorname{div} g)\|_{L_2(\Omega)}^2 + \mu \||g|\|_{L_p(\Omega)} \ ,$$

with $f \in L_2(\Omega)$, $\Omega \subset \mathbb{R}^2$, and $v = \operatorname{div} g = \operatorname{div}(g_1, g_2)$. The first term is the total variation of u. If $u \in L_1$ and $|\nabla u|$ is a finite measure on Ω, then $u \in BV(\Omega)$. This space allows discontinuities, therefore edges and contours generally appear in u. The second term represents the restoration discrepancy; to penalize v, the third term approximates (by taking p finite) the norm of the space G of oscillating functions introduced by Y. Meyer (with $p = \infty$) which is in some sense dual to $BV(\Omega)$. (For details we refer the reader to (Meyer, 2002).) Setting $p = 2$ and $g = \nabla P + Q$, where P is a single–valued function and Q is a divergence–free vector field, it is shown in (Osher et al., 2003) that the v–penalty term can be expressed by

$$\||g|\|_{L_2(\Omega)} = \left(\int_\Omega |\nabla(\Delta)^{-1} v|^2 \right)^{1/2} = \|v\|_{H^{-1}(\Omega)} \ .$$

(The H^{-1} calculus is allowed as long as we deal with oscillatory texture/noise components that have zero mean.) With these assumptions, the variational problem (2.1) simplifies to solving

$$\inf_{u,g_1,g_2} G_2(u, v) \ , \quad \text{where} \tag{2.2}$$

3

$$G_2(u,v) = \int_\Omega |\nabla u| + \lambda \|f - (u + v)\|^2_{L_2(\Omega)} + \mu \|v\|_{H^{-1}(\Omega)} \ .$$

In general, one drawback is that the minimization of (2.1) or (2.2) leads to numerically intensive schemes.

Instead of solving problem (2.2) by means of nonlinear partial differential equations and finite difference schemes, we propose a wavelet–based treatment. We are encouraged by the fact that elementary methods based on wavelet shrinkage solve similar extremal problems where $BV(\Omega)$ is replaced by the Besov space $B_1^1(L_1(\Omega))$. Since $BV(\Omega)$ can not be simply described in terms of wavelet coefficients, it is not clear that $BV(\Omega)$ minimizers can be obtained in this way. Yet, it is shown in (Cohen et al., 1999), exploiting $B_1^1(L_1(\Omega)) \subset BV(\Omega) \subset B_1^1(L_1(\Omega)) - weak$, that methods using Haar systems provide near $BV(\Omega)$ minimizers. So far there exists no similar result for general (in particular smoother) wavelet systems. We shall nevertheless use wavelets that have more smoothness/vanishing moments than Haar wavelets, because we expect them to be better suited to the modeling of the smooth parts in the cartoon image. Though we may not obtain provable 'near–best–BV–minimizers', we hope to nevertheless be 'not far off'. Limiting ourselves to the case $p = 2$, replacing $BV(\Omega)$ by $B_1^1(L_1(\Omega))$, and, moreover, extending the range of applicability by incorporating a bounded linear operator K, we end up with the following variational problem:

$$\inf_{u,v} \mathcal{F}_f(v,u) \ , \quad \text{where}$$

$$\mathcal{F}_f(v,u) = \|f - K(u + v)\|^2_{L_2(\Omega)} + \gamma \|v\|^2_{H^{-1}(\Omega)} + 2\alpha |u|_{B_1^1(L_1(\Omega))} \ . \tag{2.3}$$

In order to establish now a wavelet–based scheme that solves the latter problem, we firstly need to recall some basic facts on wavelets.

2.2 Preliminaries on wavelets

Let us briefly recall some facts on wavelets that are needed later on. Especially important for our approach are the smoothness characterization properties of wavelets: one can determine the membership of a function in many different smoothness functional spaces by examining the decay properties of its wavelets coefficients. For a comprehensive introduction and overview on this topic we would refer the reader to the abundant literature, see e.g. (Daubechies, 1992, 1993; Cohen et al., 1992; Dahmen, 1996; DeVore et al., 1992, 1988; Frazier and Jawerth, 1990; Triebel, 1978).

For readers interested more in the gist of the theory than in a more elaborate, mathematically precise description, it suffices to know that:

- wavelet expansions provide successive approximations at increasingly finer scales. If a function f is given, and $f_{\langle J \rangle}$ is its approximation at scale 2^{-J}, then the next finer

approximation $f_{\langle J+1 \rangle}$ can be written as

$$f_{\langle J+1 \rangle} = f_{\langle J \rangle} + \sum_{i,\boldsymbol{k}} \langle f, \tilde{\psi}^i_{J,\boldsymbol{k}} \rangle \psi^i_{J,\boldsymbol{k}},$$

where $\psi^i_{j,\boldsymbol{k}}(\boldsymbol{x}) = 2^j \psi^i(2^j x_1 - k_1, 2^j x_2 - k_2)$ are the wavelets used in the expansion, and $\psi^i_{J,\boldsymbol{k}}$ a corresponding dual family. The index i indicates that in dimensions larger than 1 one typically uses several wavelet templates. In 2 dimensions, there are usually 3 different wavelets, and i takes the values 1,2,3. (Note that the details of the approximation scheme that computes $f_{\langle J \rangle}$ from f depend on the wavelet family under consideration.) If $\psi \in C^s$ (i.e. ψ has 'differentiability' of order s, where s need not to be integer), then f has differentiability of order $r < s$ if and only if

$$|\langle f, \tilde{\psi}^i_{j,\boldsymbol{k}} \rangle| \le C 2^{-j(r+s)} . \tag{2.4}$$

For the sake of convenience, we shall often 'bundle' i, j, k into one index λ, and write $\langle f, \tilde{\psi}_\lambda \rangle$ simply as f_λ. In this case $|\lambda|$ stands for j. In this notation, the requirement (2.4) becomes $|f_\lambda| \le C 2^{-|\lambda|(r+s)}$.

- one can characterize the smoothness of f in detail by using several parameters to describe it, such as e.g. in Besov spaces. For smoothness $r < 1$, for instance, we define

$$\omega_l(f;t)_p = \sup_{|\boldsymbol{h}| \le t} \left[\int |f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x})|^p d\boldsymbol{x} \right]^{1/p}$$

(this is an 'L_p–measured modulus of continuity' for f), and

$$|f|_{B^r_q(L_p(\Omega))} = \left(\int_0^\infty (t^{-r} \omega(f;t)_p)^q dt/t \right)^{1/q} .$$

(Basically, this measures, in a fine 'q–gained scale', whether $\omega(f;t)_p$ decays at least as fast as t^r when $t \to 0$.) For instance, if we consider, on $\Omega = (0, 1]^2$ the function $f(\boldsymbol{x}) = x_1 + x_2 - \lfloor x_1 + x_2 \rfloor$, where $\lfloor x \rfloor = \max\{n \in \mathbb{Z}; n \le x\}$, which has a discontinuity along the diagonal $x_1 + x_2 = 1$ in the square, then we find

$$\omega(f;t)_1 \sim C|t| \quad \text{as } |t| \to 0$$

and one easily checks, $|f|_{B_1^{1-\varepsilon}(L_1(\Omega))} < \infty$ for all $\varepsilon > 0$. In fact, one has $f \in B_1^1(L_1(\Omega))$ (i.e. with $r = 1$) as well, but to verify this we need a fancier 'L_1–measured modulus of continuity'. One important link of wavelets to these detailed smoothness spaces is that they provide a good estimate of Besov norms. In particular, in 2 dimensions,

$$f \in B_1^{s+1}(L_1(\Omega)) \Leftrightarrow \sum_\lambda 2^{|\lambda|s} |f_\lambda| < \infty ;$$

for $s = 0$ this shows that $f \in B_1^1(L_1(\Omega))$ if and only if its coefficients are in ℓ_1. Another special case, again in 2 dimensions, is $p = q = 2$; the Besov spaces then reduce

5

to Sobolev spaces: $B_2^s(L_2(\Omega)) = W_2^s$, which can on Ω also be easily characterized in terms of Fourier coefficients:

$$f \in W_2^s \Leftrightarrow \sum_{\boldsymbol{k}} (|k_1| + |k_2|)^{2s} |\hat{f}_{\boldsymbol{k}}|^2 < \infty .$$

For these spaces, we have

$$f \in W_2^s = B_2^2(L_2(\Omega)) \Leftrightarrow \sum_{\lambda} 2^{2|\lambda|s} |f_\lambda|^2 < \infty .$$

This holds even for $s < 0$; in that case, functions with modest W_2^s–norm have large amplitude high frequency oscillations.

2.3 Iterative strategy for image decomposition

We aim to find the minimizer of the functional

$$\mathcal{F}_f(v, u) = \|f - K(u + v)\|_{L_2(\Omega)}^2 + \gamma \|v\|_{H^{-1}(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} . \tag{2.5}$$

At first, we may observe the following

Lemma 2.1 *If the null–space $\mathcal{N}(K)$ of the operator K is trivial, then the variational problem (2.5) has a unique minimizer.*

This can be seen as follows:

$$\mathcal{F}_f(\mu(v, u) + (1 - \mu)(v', u')) - \mu \mathcal{F}_f((v, u)) - (1 - \mu)\mathcal{F}_f((v', u')) =$$

$$-\mu(1 - \mu) \left(\|K(u - u' + v - v')\|_{L_2(\Omega)}^2 + \gamma \|v - v'\|_{H^{-1}(\Omega)}^2 \right)$$

$$+2\alpha \left(|\mu u + (1 - \mu)u'|_{B_1^1(L_1(\Omega))} - \mu |u|_{B_1^1(L_1(\Omega))} - (1 - \mu)|u'|_{B_1^1(L_1(\Omega))} \right) \tag{2.6}$$

with $0 < \mu < 1$. Since the Banach norm is convex the right hand side of (2.6) is non-positive, i.e. $\mathcal{F}_f$ is convex. Since $\mathcal{N}(K) = \{0\}$, the term $\|K(u - u' + v - v')\|$ can be zero only if $u - u' + v - v' = 0$, moreover, $\|v - v'\|$ is zero only if $v - v' = 0$. Hence, (2.6) is strictly less than zero if $(v, u) \neq (v', u')$, i.e. $\mathcal{F}_f$ is strictly convex. On the other hand, because $\mathcal{F}_f(v, u) \to \infty$ as $\|v\|, \|u\| \to \infty$, $\mathcal{F}_f$ must have a minimizer. $\qquad \square$

In order to solve this problem by means of wavelets we have to switch to the sequence space formulation. When K is the identity operator the problem simplifies to

$$\inf_{u,v} \left\{ \sum_{\lambda \in J} \left(|f_\lambda - (u_\lambda + v_\lambda)|^2 + \gamma 2^{-2|\lambda|} |v_\lambda|^2 + 2\alpha |u_\lambda| \right) \right\} , \tag{2.7}$$

where $J = \{\lambda = (i, j, k) : k \in J_j, j \in \mathbb{Z}, i = 1, 2, 3\}$ is the index set used in our separable setting. The minimization of (2.7) is straightforward, since it decouples into easy one–dimensional minimizations. This results in an explicit shrinkage scheme, presented also in (Daubechies and Teschke, 2004):

Proposition 2.1 *Let f be a given function. The functional (2.7) is minimized by the parameterized class of functions $\tilde{v}_{\gamma,\alpha}$ and $\tilde{u}_{\gamma,\alpha}$ given by the following non-linear filtering of the wavelet series of f:*

$$\tilde{v}_{\gamma,\alpha} = \sum_{\lambda \in J_{j_0}} (1 + \gamma 2^{-2|\lambda|})^{-1} \left[f_\lambda - S_{\alpha(2^{2|\lambda|}+\gamma)/\gamma}(f_\lambda) \right] \psi_\lambda$$

and

$$\tilde{u}_{\gamma,\alpha} = f_{\langle j_0 \rangle} + \sum_{\lambda \in J_{j_0}} S_{\alpha(2^{2|\lambda|}+\gamma)/\gamma}(f_\lambda) \psi_\lambda \ ,$$

where S_t denotes the soft-shrinkage operator, J_{j_0} all indices λ for scales larger than j_0 and $f_{\langle j_0 \rangle}$ is the approximation at the coarsest scale j_0.

In the case where K is not the identity operator the minimization process results in a coupled system of nonlinear equations for the wavelet coefficients u_λ and v_λ, which is not as straightforward to solve. To overcome this problem, we adapt an iterative approach. As in (Daubechies et al., 2004) we derive the iterative algorithm from a sequence of so-called surrogate functionals that are each easy to minimize, and for which one hopes that the successive minimizers have the minimizing element of (2.5) as limit. However, contrary to (Daubechies et al., 2004) our variational problem has mixed quadratic and non-quadratic penalties. This requires a slightly different use of surrogate functionals. In (Defrise and DeMol, 2004b,a) a similar $u + v$ problem is solved by an approach that combines u and v into one vector–valued function (u, v).

We will follow a different approach here, in which we first solve the quadratic problem for v, and then construct an iteration scheme for u. To this end, we introduce the differential operator $T := (-\Delta)^{1/2}$. Setting $v = Th$ the variational problem (2.5) reads as

$$\inf_{(u,h)} \mathcal{F}_f(h, u) \ , \quad \text{with} \tag{2.8}$$

$$\mathcal{F}_f(h, u) = \|f - K(u + Th)\|_{L_2(\Omega)}^2 + \gamma \|h\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} \ .$$

Minimizing (2.8) with respect to w results in

$$\tilde{h}_\gamma(f, u) = (T^* K^* K T + \gamma)^{-1} T^* K^* (f - Ku)$$

or equivalently

$$\tilde{v}_\gamma(f, u) = T(T^* K^* K T + \gamma)^{-1} T^* K^* (f - Ku) \ .$$

Inserting this explicit expression for $\tilde{h}_\gamma(f, u)$ in (2.8) and defining

$$f_\gamma := T_\gamma f, \quad T_\gamma^2 := I - KT(T^* K^* K T + \gamma)^{-1} T^* K^* \ , \tag{2.9}$$

7

we obtain

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|f_\gamma - T_\gamma K u\|^2_{L_2(\Omega)} + 2\alpha |u|_{B^1_1(L_1(\Omega))} . \tag{2.10}$$

Thus, the remaining task is to solve

$$\inf_u \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) , \quad \text{where} \tag{2.11}$$

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|f_\gamma - T_\gamma K u\|^2_{L_2(\Omega)} + 2\alpha |u|_{B^1_1(L_1(\Omega))} .$$

The corresponding variational equations in the sequence space representation are

$$\forall \lambda : \quad (K^* T_\gamma^2 K u)_\lambda - (K^* f_\gamma)_\lambda + \alpha \text{sign}(u_\lambda) = 0 .$$

This gives a coupled system of nonlinear equations for u_λ. For this reason we construct surrogate functionals that remove the influence of $K^* T_\gamma^2 K u$. First, we choose a constant C such that $\|K^* T_\gamma^2 K\| < C$. Since $\|T_\gamma\| \le 1$, it suffices to require that $\|K^* K\| < C$. Then we define the functional

$$\Phi(u; a) := C\|u - a\|^2_{L_2(\Omega)} - \|T_\gamma K(u - a)\|^2_{L_2(\Omega)}$$

which depends on an auxiliary element $a \in L_2(\Omega)$. We observe that $\Phi(u, a)$ is strictly convex in u for any a. Since K can be rescaled, we limit our analysis without loss of generality to the case $C = 1$. We finally add $\Phi(u; a)$ to $\mathcal{F}_f(\tilde{h}_\gamma(f, u), u)$ and obtain the following surrogate functional

$$\begin{aligned}
\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, a), u; a) &= \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) + \Phi(u; a) \\
&= \sum_\lambda \{u_\lambda^2 - 2u_\lambda(a + K^* T_\gamma^2(f - Ka))_\lambda + 2\alpha |u_\lambda|\} \\
&\quad + \|f_\gamma\|^2_{L_2(\Omega)} + \|a\|^2_{L_2(\Omega)} - \|T_\gamma Ka\|^2_{L_2(\Omega)} . \tag{2.12}
\end{aligned}$$

The advantage of minimizing (2.12) is that the variational equations for u_λ decouple. The summands of (2.12) are differentiable in u_λ except at the point of non-differentiability. The variational equations for each λ are now given by

$$u_\lambda + \alpha \text{sign}(u_\lambda) = (a + K^* T_\gamma^2(f - Ka))_\lambda .$$

This results in an explicit soft-shrinkage operation for u_λ

$$u_\lambda = S_\alpha((a + K^* T_\gamma^2(f - Ka))_\lambda) .$$

The next proposition summarizes our findings; it is the specialization to our particular case of a more general theorem in (Daubechies et al., 2004).

Proposition 2.2 *Suppose K is a linear bounded operator modeling the blur, with K maps $L_2(\Omega)$ to $L_2(\Omega)$ and $\|K^* K\| < 1$. Moreover, assume T_γ is defined as in (2.9) and the functional $\mathcal{F}_f^{sur}(h, u; a)$ is given by*

$$\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; a) = \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) + \Phi(u; a) .$$

Then, for arbitrarily chosen $a \in L_2(\Omega)$, the functional $\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f,u), u; a)$ has a unique minimizer in $L_2(\Omega)$. The minimizing element is given by

$$\tilde{u}_{\gamma,\alpha} = \mathbf{S}_\alpha(a + K^* T_\gamma^2(f - Ka)) \ ,$$

where the operator $\mathbf{S}_\alpha$ is defined component-wise by

$$\mathbf{S}_\alpha(x) = \sum_\lambda S_\alpha(x_\lambda) \psi_\lambda \ .$$

The proof follows from (Daubechies et al., 2004). One can now define an iterative algorithm by repeated minimization of $\mathcal{F}_f^{sur}$:

$$u^0 \text{ arbitrary} \ ; \quad u^n = \arg\min_u \left(\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f,u), u; u^{n-1}) \right) \quad n = 1, 2, \ldots \tag{2.13}$$

The convergence result of (Daubechies et al., 2004) can again be applied directly:

Theorem 2.1 *Suppose K is a linear bounded operator, with $\|K^*K\| < 1$, and that T_γ is defined as in (2.9). Then the sequence of iterates*

$$u_{\gamma,\alpha}^n = \mathbf{S}_\alpha(u_{\gamma,\alpha}^{n-1} + K^* T_\gamma^2(f - K u_{\gamma,\alpha}^{n-1})) \ , \quad n = 1, 2, \ldots \quad ,$$

with arbitrarily chosen $u^0 \in L_2(\Omega)$, converges in norm to a minimizer $\tilde{u}_{\gamma,\alpha}$ of the functional

$$\mathcal{F}_f(\tilde{h}_\gamma(f,u), u) = \|T_\gamma(f - Ku)\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} \ .$$

If $\mathcal{N}(T_\gamma K) = \{0\}$, then the minimizer $\tilde{u}_{\gamma,\alpha}$ is unique, and every sequence of iterates converges to $\tilde{u}_{\gamma,\alpha}$ in norm.

Combining the result of Theorem 2.1 and the representation for $\tilde{v}$ we summarize how the image can finally be decomposed in cartoon and oscillating components.

Corollary 2.1 *Assume that K is a linear bounded operator modeling the blur, with $\|K^*K\| < 1$. Moreover, if T_γ is defined as in (2.9) and if $\tilde{u}_{\gamma,\alpha}$ is the minimizing element of (2.11), obtained as a limit of $u_{\gamma,\alpha}^n$ (see Theorem 2.1), then the variational problem*

$$\inf_{(u,h)} \mathcal{F}_f(h, u), \text{ with } \mathcal{F}_f(h, u) = \|f - K(u + Th)\|_{L_2(\Omega)}^2 + \gamma \|h\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))}$$

is minimized by the class

$$\left(\tilde{u}_{\gamma,\alpha}, (T^* K^* K T + \gamma)^{-1} T^* K^* (f - K \tilde{u}_{\gamma,\alpha}) \right) \ ,$$

where $\tilde{u}_{\gamma,\alpha}$ is the unique limit of the sequence

$$u_{\gamma,\alpha}^n = \mathbf{S}_\alpha(u_{\gamma,\alpha}^{n-1} + K^* T_\gamma^2(f - K u_{\gamma,\alpha}^{n-1})) \ , \quad n = 1, 2, \ldots \quad .$$
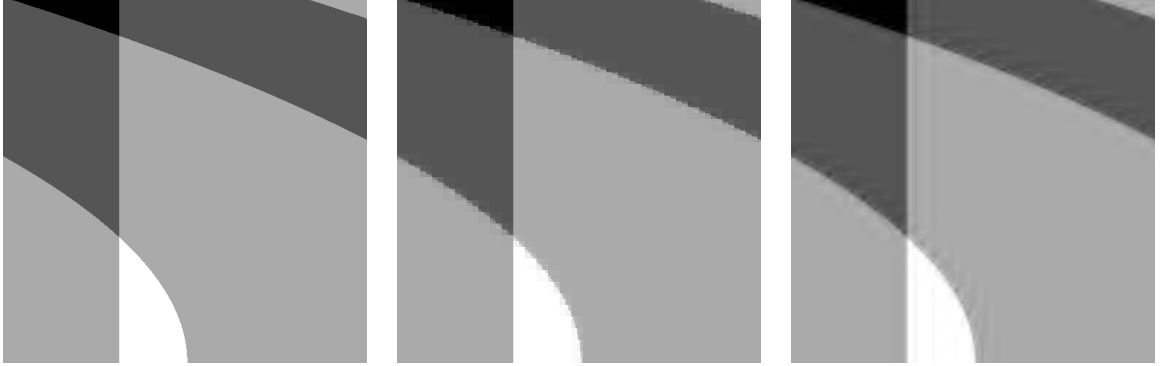
9

Figure 1: An initial geometric image f (left), and two versions of f (the middle decomposed with the Haar wavelet basis and the right with the Db3 basis) where the soft-shrinkage operator with shrinkage parameter $\alpha = 0.5$ was applied.

2.4   Redundancy and adaptivity to reduce artifacts

The non-linear filtering rule of Proposition 2.1 gives explicit descriptions of $\tilde{v}$ and $\tilde{u}$ that are computed by fast discrete wavelet schemes. However, non-redundant filtering very often creates artifacts in terms of undesirable oscillations, which manifest themselves as ringing and edge blurring, see Figure 1. Poor directional selectivity of traditional tensor product wavelet bases likewise cause artifacts. In this section we discuss various refinements on the basic algorithm that address this problem. In particular, we shall use redundant translation invariant schemes, complex wavelets, and additional edge dependent penalty weights. We describe these generalizations here, and leave examples to the next section.

2.4.1   Translation invariance by cycle–spinning

Assume that we are given an image with 2^M rows of 2^M pixels, where the gray value of each pixel gives an average of f on a square $2^{-M} \times 2^{-M}$, which we denote by f_k^M, with k a double index running through all the elements of $\{0, 1, \ldots, 2^M - 1\} \times \{0, 1, \ldots, 2^M - 1\}$. A traditional wavelet transform then computes $f_{\langle j_0 \rangle}$ and $f_{j,\boldsymbol{l},i}$ with $j_0 \leq j \leq M$, $i = 1, 2, 3$ and $\boldsymbol{l} \in \{0, 1, \ldots, 2^j - 1\} \times \{0, 1, \ldots, 2^j - 1\}$ for each j, where the $f_{j,\boldsymbol{l},i}$ stand for the different species of wavelet coefficients (in two dimensions, there are three), mostly localized on (and indexed by) the squares $[l_1 2^{-j}, (l_1 + 1)2^{-j}] \times [l_2 2^{-j}, (l_2 + 1)2^{-j}]$.

Because the corresponding wavelet basis is not translation invariant (as can be seen from the localization of the wavelet coefficients; $f_{\langle j_0 \rangle}$ has a similar translation non–invariance which we did not denote explicitly), Coifman and Donoho proposed in (Coifman and Donoho, 1995) to recover translation invariance by averaging over the $2^{2(M+1-j_0)}$ translates of the wavelet basis; since many wavelets occur in more than one of these

translated bases (in fact, each $\psi^i_{j,k}(x - 2^M \boldsymbol{n})$ in exactly $2^{2(j+1-j_0)}$ different bases), the average over all these bases uses only $(M + 1 - j_0)2^{2M}$ different basis functions (and not $2^{4(M+1-j_0)} = $ number of bases × number of elements in each basis). This approach is called *cycle–spinning*. Writing, with a slight abuse of notation, $\psi^i_{j,\boldsymbol{k}+2^{j-M}\boldsymbol{n}}$ for the translate $\psi^i_{j,\boldsymbol{k}}(x - 2^M \boldsymbol{n})$, this average can then be written as

$$f^M = f^{\text{cycled}}_{\langle j_0 \rangle} + 2^{-2(M+1)} \sum_{l_1,l_2=0}^{2^M-1} \sum_{j=j_0}^{M-1} 2^{2j} \sum_{i=1}^{3} f^i_{j,\boldsymbol{l}2^{-M+j}} \psi^i_{j,\boldsymbol{l}2^{-M+j}} \ .$$

Carrying out our nonlinear filtering in each of the bases and averaging the result then corresponds to applying the corresponding nonlinear filtering on the (much smaller number of) coefficients in the last expression. This is the standard way to implement thresholding on cycle–spinned representations.

The resulting sequence space representation of the variational functional (2.7) has to be adapted to the redundant representation of f. To this end, we note that the Besov penalty term takes the form

$$|f|_{B^s_p(L_p)} \sim \left(\sum_{|\lambda| \geq j_0} 2^{(j(s-1+2/p)+2(j-M))p} |f^i_{j,\boldsymbol{l}2^{-M+j}}|^p \right)^{1/p} \ .$$

The norms $\| \cdot \|^2_{L_2}$ and $\| \cdot \|^2_{H^{-1}}$ change similarly. Consequently, we obtain the same minimization rule but with respect to a richer class of wavelet coefficients.

2.4.2 Directional sensitivity by frequency projections

It has been shown by several authors (Kinsbury, 1999; Selesnick, 2001; Fernandes et al., 2000) that if one treats positive and negative frequencies separately in the one–dimensional wavelet transform (resulting in complex wavelets), the directional selectivity of the corresponding two–dimensional multi–resolution analysis is improved. This can be done by applying the following orthogonal projections:

$$\begin{aligned} \mathcal{P}^+ &: \quad L_2 \to L_{2,+} = \{f \in L_2 : \ \text{supp } \hat{f} \subseteq [0, \infty)\} \\ \mathcal{P}^- &: \quad L_2 \to L_{2,-} = \{f \in L_2 : \ \text{supp } \hat{f} \subseteq (-\infty, 0]\} \ . \end{aligned}$$

The projectors $\mathcal{P}^+$ and $\mathcal{P}^-$ may be either applied to f or to $\{\phi, \tilde{\phi}\}$ and $\{\psi, \tilde{\psi}\}$. In a discrete framework these projections have to be approximated. This has been done in different ways in the literature. In (Kinsbury, 1999; Selesnick, 2001) Hilbert transform pairs of wavelets are used. In (Fernandes et al., 2000) f is projected (approximately) by multiplying with shifted generator symbols in the frequency domain. We follow the second approach, i.e.

$$(P^+ f)^\wedge(\omega) := \hat{f}(\omega)H(\omega - \pi/2) \ \text{ and } \ (P^- f)^\wedge(\omega) := \hat{f}(\omega)H(\omega + \pi/2) \ ,$$

where f denotes the function to be analyzed and H is the low–pass filter for a conjugate quadrature mirror filter pair. One then has

$$\hat{f}(\omega) = (B^+ P^+ f)^\wedge(\omega) + (B^- P^- f)^\wedge(\omega) , \qquad (2.14)$$

where the back–projections are given by

$$(B^+ f)^\wedge = \hat{f}\,\overline{H(\cdot - \pi/2)} \quad \text{and} \quad (B^- f)^\wedge = \hat{f}\,\overline{H(\cdot + \pi/2)}$$

respectively. This technique provides us with a simple multiplication scheme in Fourier, or equivalently, a convolution scheme in time domain. In a separable two–dimensional framework the projections need to be carried out in each of the two frequency variables, resulting in four approximate projection operators $P^{++}, P^{+-}, P^{-+}, P^{--}$. Because f is real, we have

$$(P^{++} f)^\wedge(-\omega) = \overline{(P^{--} f)^\wedge(\omega)} \quad \text{and} \quad (P^{+-} f)^\wedge(-\omega) = \overline{(P^{-+} f)^\wedge(\omega)} ,$$

so that the computation of $P^{-+} f$ and $P^{--} f$ can be omitted. Consequently, the modified variational functional takes the form

$$
\begin{aligned}
\mathcal{F}_f(u,v) \;=\; & 2 \left(\| P^{++}(f - (u+v)) \|_{L_2}^2 + \| P^{+-}(f - (u+v)) \|_{L_2}^2 \right) + \\[2mm]
& 2\lambda \left(\| P^{++} v \|_{H^{-1}}^2 + \| P^{+-} v \|_{H^{-1}}^2 \right) + 2\alpha |u|_{B^1_{1(L_1)}} \\[2mm]
\;\leq\; & 2 \left(\| P^{++}(f - (u+v)) \|_{L_2}^2 + \| P^{+-}(f - (u+v)) \|_{L_2}^2 \right) + \\[2mm]
& 2\lambda \left(\| P^{++} v \|_{H^{-1}}^2 + \| P^{+-} v \|_{H^{-1}}^2 \right) + \\[2mm]
& 4\alpha \left(|P^{++} u|_{B^1_{1(L_1)}} + |P^{+-} u|_{B^1_{1(L_1)}} \right) ,
\end{aligned}
$$

which can be minimized with respect to $\{P^{++} v, P^{++} u\}$ and $\{P^{+-} v, P^{+-} u\}$ separately. The projections are complex–valued, so that the thresholding operator needs to be adapted. Parameterizing the wavelet coefficients by modulus and angle and minimizing yields the following filtering rules for the projections of $\tilde{v}_{\gamma,\alpha}$ and $\tilde{u}_{\gamma,\alpha}$ (where $\cdot\cdot$ stands for any combination of $+, -$)

$$P^{\cdot\cdot} \tilde{v}_{\gamma,\alpha} = \sum_{|\lambda| \geq j_0} (1 + \gamma 2^{-2|\lambda|})^{-1} \left[P^{\cdot\cdot} f_\lambda - S_{\alpha(2^{2|\lambda|}+\gamma)/\gamma}(|P^{\cdot\cdot} f_\lambda|) e^{i\omega(P^{\cdot\cdot} f)} \right] \psi_\lambda$$

and

$$P^{\cdot\cdot} \tilde{u}_{\gamma,\alpha} = (P^{\cdot\cdot} f)_{\langle j_0 \rangle} + \sum_{|\lambda| \geq j_0} (1 + \gamma 2^{-2|\lambda|})^{-1} S_{\alpha(2^{2|\lambda|}+\gamma)/\gamma}(|P^{\cdot\cdot} f_\lambda|) e^{i\omega(P^{\cdot\cdot} f)} \psi_\lambda .$$

Finally, we have to apply the back-projections to obtain the minimizing functions

$$\tilde{v}_{\gamma,\alpha}^{BP} = B^{++} P^{++} \tilde{v}_{\gamma,\alpha} + B^{--} \overline{P^{++} \tilde{v}_{\gamma,\alpha}} + B^{+-} P^{+-} \tilde{v}_{\gamma,\alpha} + B^{-+} \overline{P^{+-} \tilde{v}_{\gamma,\alpha}}$$

and

$$\tilde{u}_{\gamma,\alpha}^{BP} = B^{++} P^{++} \tilde{u}_{\gamma,\alpha} + B^{--} \overline{P^{++} \tilde{u}_{\gamma,\alpha}} + B^{+-} P^{+-} \tilde{u}_{\gamma,\alpha} + B^{-+} \overline{P^{+-} \tilde{u}_{\gamma,\alpha}} .$$

2.4.3 Weighted penalty functions

In order to improve the capability of preserving edges we additionally introduce a positive weight sequence w_λ in the H^{-1} penalty term. Consequently, we aim at minimizing a slightly modified sequence space functional

$$\sum_\lambda \left(|f_\lambda - (u_\lambda + v_\lambda)|^2 + \gamma 2^{-2|\lambda|} w_\lambda |v_\lambda|^2 + 2\alpha |u_\lambda| \right) . \tag{2.15}$$

The resulting texture and cartoon components take the form

$$\tilde{v}^w_{\gamma,\alpha} = \sum_{|\lambda| \geq j_0} (1 + \gamma w_\lambda 2^{-2|\lambda|})^{-1} \left[f_\lambda - S_{\alpha(2^{2|\lambda|} + \gamma w_\lambda)/\gamma w_\lambda}(f_\lambda) \right] \psi_\lambda$$

and

$$\tilde{u}^w_{\gamma,\alpha} = f_{\langle j_0 \rangle} + \sum_{|\lambda| \geq j_0} S_{\alpha(2^{2|\lambda|} + \gamma w_\lambda)/\gamma w_\lambda}(f_\lambda) \psi_\lambda .$$

The main goal is to introduce a control parameter that depends on the local structure of f. The local penalty weight w_λ should be large in the presence of an edge and small otherwise; the result of this weighting is to enhance the sensitivity of u near edges. In order to do this, we must first localize the edges, which we do by a procedure similar to an edge detection algorithm in (Mallat and Zhong, 1992). This scheme rests on the analysis of the cycle-spinned wavelet coefficients f_λ at or near the same location but at different scales. We expect that the f_λ belonging to fine decomposition scales contain informations of edges (well localized) as well as oscillating components. Oscillating texture components typically show up in fine scales only; edges on the other hand leave a signature of larger wavelet coefficients through a wider range of scales. We thus apply the following not very sophisticated edge detector. Suppose that $f \in V_M$ and j_e denotes some 'critical' scale, then for a certain range of scales $|\lambda| = |(i, j, k)| = j \in \{j_0, \ldots, j_1 - j_e - 2, j_1 - j_e - 1\}$ we mark all positions k where $|f_\lambda|$ is larger than a level dependent threshold parameter t_j. Here the value t_j is chosen proportional to the mean value of all wavelet coefficients of level j. We say that $|f_\lambda|$ represents an edge if k was marked for all $j \in \{j_0, \ldots, j_1 - j_e - 2, j_1 - j_e - 1\}$. Finally, we adaptively choose the penalty sequence by setting

$$w_\lambda = \begin{cases} \Theta_\lambda & \text{if } j \in \{M - 1, \ldots, j_1 - j_e\} \text{ and } k \text{ was marked as an edge ,} \\ \vartheta_\lambda & \text{otherwise ,} \end{cases}$$

where ϑ_λ is close to one and Θ_λ is much larger in order to penalize the corresponding v_λ's.

2.5 Image examples

In this section, we present some examples of images decomposed, deblurred and denoised.

We start with the case where K is the identity operator. In order to show how the nonlinear (redundant) wavelet scheme acts on piecewise constant functions we decompose
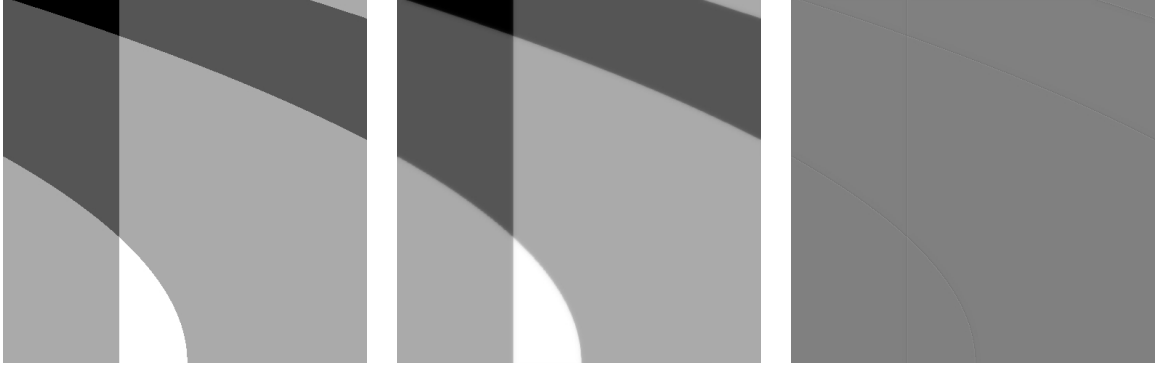
Figure 2: From left to right: initial geometric image f, $\tilde{u}$, $\tilde{v} + 150$, computed with Db3 in the translation invariant setting, $\alpha = 0.5$, $\gamma = 0.01$.

Figure 3: Left: noisy segment of a woman image, middle and right: first two scales of $\mathbf{S}(f)$ inducing the weight function w.

a geometric image (representing cartoon components only) with sharp contours, see Figure 2. We observe that $\tilde{u}$ represents the cartoon part very well. The texture component $\tilde{v}$ (plus a constant for illustration purposes) contains only some very weak contour structures.

Next, we demonstrate the performance of the Haar shrinkage algorithm successively incorporating redundancy and local penalty weights. The redundancy is implemented by cycle spinning as described in Section 2.4.1. The local penalty weights are computed in the following way: first, we apply the shrinkage operator $\mathbf{S}$ to f with a level dependent threshold (the threshold per scale is equal to two times the mean value of all the wavelet coefficients of the scale under consideration). Second, for those λ according to the non–zero values of $S_{\text{threshold}}(f_\lambda)$ we set w_λ to $\Theta_\lambda = 1 + C'$ (here $C' = 10$, moreover, we set w_λ equal to $\vartheta_\lambda = 1$ elsewhere). The coefficients $S_{\text{threshold}}(f_\lambda)$ for the first two scales of a segment of the image 'Barbara' are visualized in Figure 3. In Figure 4, we present our numerical results. The upper row shows the original and the noisy image. The next row visualizes the results for non-redundant Haar shrinkage (Method A). The third row shows the same but incorporating cycle spinning (Method B), and the last row shows the

Haar Shrinkage	SNR(f, f_ε)	SNR(f,$u + v$)	SNR(f,u)
Method A	20,7203	18,3319	16,0680
Method B	20,7203	21,6672	16,5886
Method C	20,7203	23,8334	17,5070

Table 1: Signal–to–noise ratios of the several decomposition methods (Haar shrinkage, translation invariant Haar shrinkage, translation invariant Haar shrinkage with edge enhancement).

Data basis	"Barbara" image (512x512 pixel)
Hardware Architecture	PC
Operating System	linux
OS Distribution	redhat7.3
Model	PC, AMD Athlon-XP
Memory Size (MB)	1024
Processor Speed (MHz)	1333
Number of CPUs	1
Computational cost	(average over 10 runs)
PDE scheme in Fortran (compiler f77)	56,67 sec
wavelet shrinkage Method A (Matlab)	4,20 sec
wavelet shrinkage Method B (Matlab)	24,78 sec
wavelet shrinkage Method C (Matlab)	26,56 sec

Table 2: Comparison of computational cost of the PDE– and the wavelet–based methods.

incorporation of cycle spinning and local penalty weights. Each extension of the shrinkage method improves the results. This is also confirmed by comparing the signal–to–noise-ratios (which is here defined as follows: $SNR(f, g) = 10 \log_{10}(\|f\|^2/\|f - g\|^2)$), see Table 1.

The next experiment is done on a fabric image, see Figure 5. But in contrast to the examples before, we present here the use of frequency projection as introduced in Section 2.4.2. The numerical result shows convincingly that the texture component can be also well separated from the cartoon part.

In order to compare the performance with the $BV - L_2$ model (Rudin et al., 1992) and with the $BV - H^{-1}$ model (Osher et al., 2003), we apply our scheme to a woman image (the same that was used in (Vese and Osher, 2003; Osher et al., 2003)), see Figure 6. We obtain very similar results as obtained with the model proposed in (Osher et al., 2003). Compared with the results obtained with the $BV - L_2$ model (Rudin et al., 1992) we observe that our reconstruction of the texture component contains much less cartoon information. In terms of computational cost we have observed that even in the case of applying cycle spinning and edge enhancement our proposed wavelet shrinkage scheme is

15

less time consuming than the $BV - H^{-1}$ restoration scheme, see table 2, even when the wavelet method is implemented in Matlab, which is slower than the compiled version for the (Osher et al., 2003) scheme.

We end this section with an experiment where K is not the identity operator. In our particular case K is a convolution operator with Gaussian kernel. The implementation is simply done in Fourier space. The upper row in Figure 7 shows the original f and the blurred image Kf. The lower row visualizes the results: the cartoon component $\tilde{u}$, the texture component $\tilde{v}$, and the sum of both $\tilde{u} + \tilde{v}$. One may clearly see that the deblurred image $\tilde{u} + \tilde{v}$ contains (after a small number of iterations) more small scale details than Kf. This definitely shows the capabilities of the proposed iterative deblurring scheme (2.13).
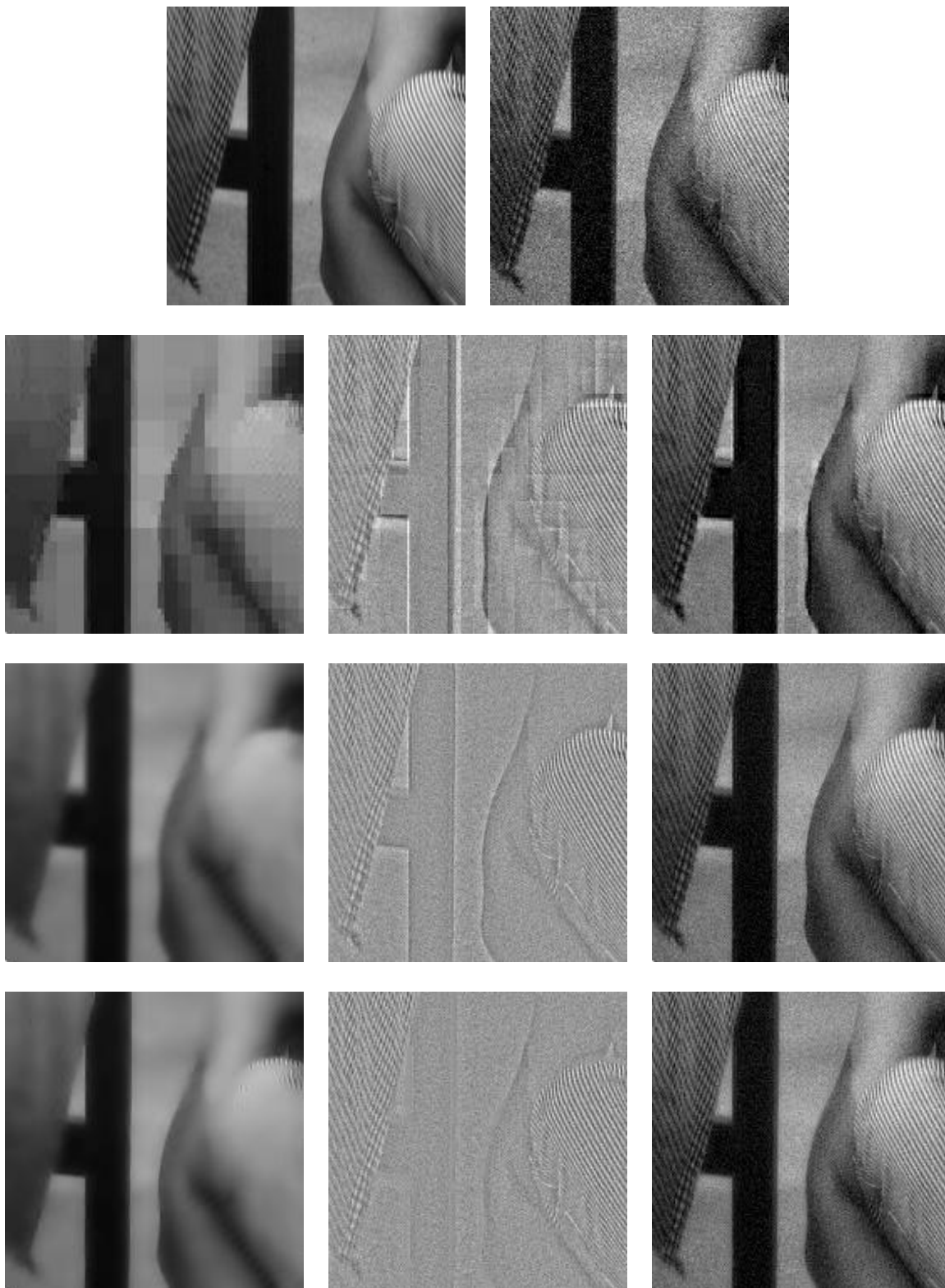
Figure 4: Top: initial and noisy image, 2nd row: non-redundant Haar shrinkage (Method A), 3rd row: translation invariant Haar shrinkage (Method B), bottom: translation invariant Haar shrinkage with edge enhancement (Method C); 2nd-4th row from left to right: $\tilde{u}$, $\tilde{v} + 150$ and $\tilde{u} + \tilde{v}$, $\alpha = 0.5$, $\gamma = 0.0001$, computed with Haar wavelets and critical scale $j_e = -3$.
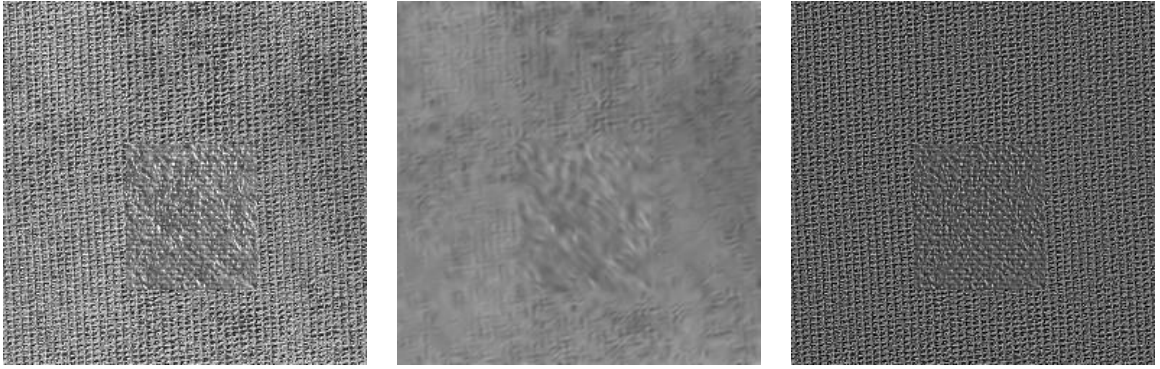
17

Figure 5: From left to right: initial fabric image f, $\tilde{u}$, $\tilde{v} + 150$, computed with Db4 incorporating frequency projections, $\alpha = 0.8$, $\gamma = 0.002$.

Figure 6: Top from left to right: initial woman image f, $\tilde{u}$ and $\tilde{v} + 150$, computed with Db10 (Method C), $\alpha = 0.5$, $\gamma = 0.002$; bottom from left to right: u and v obtained by the $BV - H^{-1}$ model (Osher et al., 2003) and the v component obtained by the classical TV model (Rudin et al., 1992).
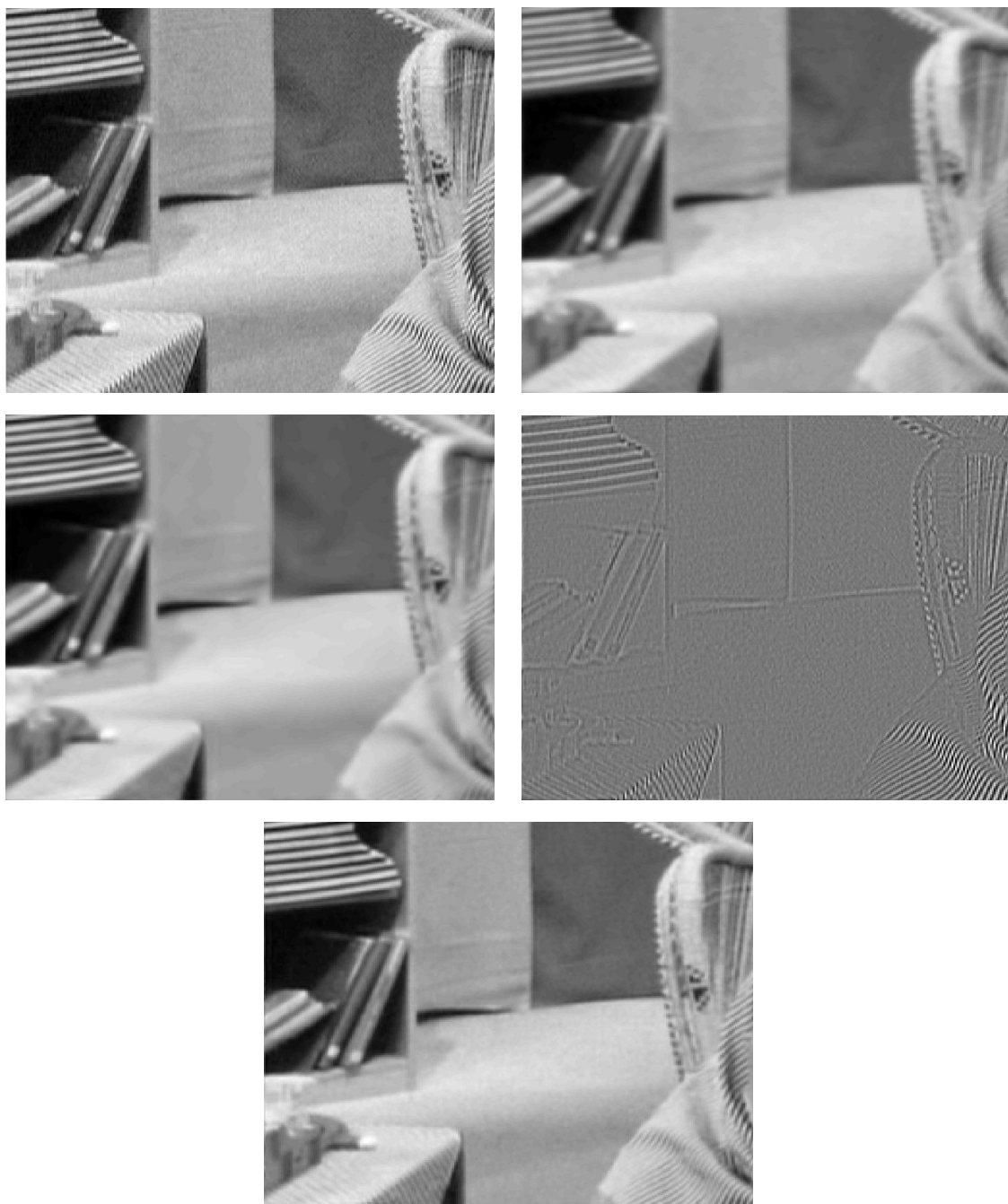
Figure 7: Top from left to right: initial image f, blurred image Kf; middle from left to right: deblurred $\tilde{u}$, deblurred $\tilde{v} + 150$; bottom: deblurred $\tilde{u} + \tilde{v}$, computed with Db3 using the iterative approach, $\alpha = 0.2$, $\gamma = 0.001$.

3 Vector-valued regimes and mixed constraints

In the previous Section 2 we have considered a concrete image decomposition problem where the solution was assumed to be a vector of functions, namely (v, u). Since the constraint on v was a quadratic one, we were able to derive for the component v an explicit expression. Therefore, problem (2.5) could be transformed into the much simpler form (2.11) for which we could directly apply the basic algorithm of (Daubechies et al., 2004).

It is now quite natural to generalize the iterative approach proposed in (Daubechies et al., 2004) to the vector–valued situation, i.e. we now assume to have an m–dimensional data vector $(f_1, \ldots, f_m)$ available from which we wish to reconstruct an n–dimensional object $(v_1, \ldots, v_n)$ and where, moreover, the constraints on the object might be a mixture of smoothness and sparsity measures. Similar problems were discussed in (Defrise and DeMol, 2004a; Anthoine, 2005; Fornasier and Rauhut, 2006; Elad et al., 2005; Starck et al., 2005). We limit ourselves here to the special case $m = 1$, with the extra assumption that $(v_1, \ldots, v_n)$ has a sparse expansion (or satisfies some other constraint) with respect to several bases or frames; the main difference with the preceding section is that we we provide a rich dictionary of bases/frames that serves as a reservoir of building blocks for $(v_1, \ldots, v_n)$. Our main motivation for this work was an approach in audio data coding by B. Torrésani et.al. (Molla and Torresani, 2005; Jaillet and Torresani, 2005; Daudet and Torrsani, 2002), who represented audio signals by means of wavelets for transients and local cosine functions for tonal components. Their approach produces sparse representations of audio signals that are very efficient in audio coding. We shall illustrate at the end of this section how the scheme developed here works for such audio coding.

3.1 Some remarks on frame dictionaries and sparsity

Sparsity can be achieved by using a suitable basis in the underlying function space. In the preceding section, we introduced redundant systems to reduce artifacts. However, recent studies indicate that redundant systems, such as frames, or dictionaries of 'waveform' systems may also lead to better, i.e. sparser representations. When dealing with dictionaries of 'waveform' systems, there exist several methods, e.g. best orthogonal basis, matching pursuit, basis pursuit etc., see, e.g., (Chen et al., 1999), that allow a decomposition of a signal into an 'optimal' superposition of dictionary elements, where optimal means having the smallest ℓ_1 norm of coefficients among all such decompositions. Numerical schemes to implement these iterative 'pursuit' schemes in highly overcomplete dictionaries often lead to very large scale optimization problems.

As an alternative to these methods, we discuss a method for finding the ℓ_p–optimal decomposition $(1 \leq p \leq 2)$ for which the skeletal structure is taken from the iterative strategy proposed in (Daubechies et al., 2004). The advantage of the discussed method

is that to achieve convergence of the iteration process, we do not need to make strong assumptions on the preselected family of frames.

3.1.1 Frames, sparsity and inverse problems

A frame $\{\phi_\lambda : \lambda \in \Lambda\}$ in a Hilbert space $\mathcal{H}$ is a set of vectors for which there exist constants $0 < A \leq B < \infty$ such that, for all $v \in \mathcal{H}$,

$$A\|v\|_{\mathcal{H}}^2 \leq \sum_{\lambda \in \Lambda} |\langle v, \phi_\lambda \rangle_{\mathcal{H}}|^2 \leq B\|v\|_{\mathcal{H}}^2 \ .$$

Frames are typically 'overcomplete', i.e. for a given vector $v \in \mathcal{H}$, one can find many different sequences $g \in \ell_2$ of coefficients so that

$$v = \sum_{\lambda \in \Lambda} g_\lambda \phi_\lambda. \tag{3.1}$$

Some of these sequences have special properties, for instance, one may prefer the sequence with minimal ℓ_2 norm. The problem of finding sequences g can be considered as an inverse problem. To this end, let us introduce the operator F (often called the frame operator) that maps a function $v \in \mathcal{H}$ to the element Fv of ℓ_2 by $Fv = \{\langle v, \phi_\lambda \rangle_{\mathcal{H}}\}_{\lambda \in \Lambda}$. The adjoint F^* maps a sequence $g \in \ell_2$ to the element F^*g of $\mathcal{H}$ via $F^*g = \sum_{\lambda \in \Lambda} g_\lambda \phi_\lambda$, i.e. solving (3.1) amounts to solving $F^*g = v$.

The sequence g with minimal ℓ_2–norm is obtained by standard least–squares methods for these equation. It is often of interest to find sequences that are sparser than the minimum ℓ_2–norm solution. For instance, if the object v is known to be a (noisy version of a) sparse linear combination of the ϕ_λ, it makes sense to seek a coefficient sequence with small ℓ_p-norm (e.g. $p = 1$), see (Daubechies et al., 2004). It then makes sense to compute the sequence g that minimizes

$$\|v - F^*g\|_{\mathcal{H}}^2 + \alpha\|g\|_{\ell_p}^p \ . \tag{3.2}$$

In many applications, the features or signals of interest cannot be observed directly, but have to be inferred from other, observable quantities. Very often, there is a linear relationship $K : \mathcal{H} \mapsto \mathcal{H}'$ between the feature, modelled by a function v, and the derived quantities, modelled by another function z, which often has additional noise; the relation between v and z can then be written as

$$f = z + e = Kv + e \ .$$

To find an estimate for v from observed f, one can minimize the discrepancy

$$\|f - Kv\|_{\mathcal{H}'}^2 \ . \tag{3.3}$$

Combining (3.2) and (3.3), we end up with

$$\Phi(g) = \|f - AF^*g\|_{\mathcal{H}'}^2 + \alpha\|g\|_{\ell_p}^p \ , \tag{3.4}$$

where we allow $1 \leq p \leq 2$. As we have seen in the previous Section 2, for this variational problem an iterative method to approximate the minimizer was suggested (Daubechies et al., 2004).

3.1.2 Extension to frame dictionaries

Instead of using one single frame only, we aim now to represent the function we are searching for by several different frames. This makes sense since there are certain classes of signals where one particular frame (or basis) is not optimally suited (in the sense of locally best sparse approximation). Since a finite union of frames is again a frame, variational formulation (3.4) applies here as well. But when putting mixed or different constraints on the different frames, a setup where each frame is treated individually is better suited. An extensive discussion on this subject can be found, e.g. in (Teschke, 2007).

We denote with $\{\phi_\lambda^i : \lambda \in \Lambda_i, i = 1, 2, \ldots, n\}$ the finite family (or dictionary) of frames where each individual collection $\{\phi_\lambda^i : \lambda \in \Lambda_i\}$ is assumed to be a frame for $\mathcal{H}$. For each frame, we may consider the associated frame operator $F_i : \mathcal{H} \to \ell_2$ which is defined by the map $v \mapsto v^i := \{\langle v, \phi_\lambda^i \rangle\}_{\lambda \in \Lambda_i}$. A natural composition of all frame operators is given by the sum of its adjoints,

$$(v^1, \ldots, v^n) \mapsto \sum_{i=1}^n F_i^* v^i \ .$$

Involving our linear relationship $K : \mathcal{H} \to \mathcal{H}'$, we may define the operator $K^n : (\ell_2)^n \to \mathcal{H}'$ by

$$K^n : (\ell_2)^n : (v^1, \ldots, v^n) \mapsto \sum_{i=1}^n K F_i^* v^i$$

where the adjoint $(K^n)^* : \mathcal{H}' \mapsto (\ell_2)^n$ is given by

$$g \mapsto K_A^* g = (F_1 A^* g, \ldots, F_n A^* g) \ .$$

With this specific operator K^n we may define the following variational functional

$$\Phi(g) := \|f - K^n g\|_{\mathcal{H}'}^2 + \alpha \cdot |||g||| \ , \tag{3.5}$$

where $g = (g^1, \ldots, g^n) \in (\ell_2)^n$, $|||g||| := (\|g^1\|_{p_1}^{p_1}, \ldots, \|g^n\|_{p_n}^{p_n})$ and $\alpha = (\alpha_1, \ldots, \alpha_n)$ represents n positive regularization parameters. As before, we restrict ourselves to $1 \leq p_i \leq 2$, but not necessarily requiring $p_i = p_j$. For $n = 2$, several concepts to minimize (3.5) are suggested in (Daubechies and Teschke, 2004, 2005; Defrise and DeMol, 2004a). In what follows, we adapt the strategy that was proposed in (Daubechies et al., 2004).

3.2 Iterative approach by surrogate functionals

At first, one easily verifies that Φ as defined in (3.5) is convex. In order to apply the technique of Gaussian surrogate functionals, we define a constant $C := \tilde{C}\sqrt{B_1 + \ldots + B_n}$, where $\tilde{C}$ is an upper bound for K and B_i stands for the upper frame bound with respect to F_i. Then, for some auxiliary element $a \in (\ell_2)^n$, the Gaussian surrogate extension for the data misfit term takes the form

$$\Gamma^{sur}(g; a) = \|f - K^n g\|_{\mathcal{H}'}^2 + C^2 \|g - a\|_{(\ell_2)^n}^2 - \|K^n g - K^n a\|_{\mathcal{H}'}^2.$$

This functional is again convex and it holds $\Gamma^{sur}(g; a) - \|f - K_A g\|_{\mathcal{H}'}^2 \geq 0$. Therefore, it is reasonable to consider instead of Φ the surrogate functional

$$\Phi^{sur}(g; a) := \Gamma^{sur}(g, a) + \alpha \cdot |\|g\|| \,, \tag{3.6}$$

satisfying for all $a \in (\ell_2)^n$, $\Phi^{sur}(g; g) = \Phi(g)$ and $\Phi^{sur}(g; a) \geq \Phi(g)$. To approach the minimizer $g^\star$ of (3.5), we consider the following iteration:

$$g_0 \text{ arbitrary} \,; \quad g_{m+1} = \arg\min_g \Phi^{sur}(g; g_m) \quad m = 0, 1, \ldots \tag{3.7}$$

In order to execute iteration (3.7), we have to evaluate the necessary conditions for a minimum of (3.6). For some generic $a \in (\ell_2)^n$ we have

$$\Phi^{sur}(g; a) = \sum_{i=1}^{n} \sum_{\lambda \in \Lambda_i} \left(C^2 (g_\lambda^i)^2 - 2g_\lambda^i \left[F_i A^* f + C^2 a^i - F_i K^* K^n a \right]_\lambda + \alpha_i |g_\lambda^i|^{p_i} \right)$$

$$+ \|f\|_{\mathcal{H}'}^2 + C^2 \|a\|_{(\ell_2)^n}^2 - \|K^n a\|_{\mathcal{H}'}^2 \,, \tag{3.8}$$

where g_λ^i stands for the coefficients of g^i. We observe that through the Gaussian surrogate extension the variational equations for the individual g_λ^i decouple which allows - as we shall see - an explicit computation of the minimizer.

Each summand in (3.8) is differentiable in g_λ^i except for $p_i = 1$ at $g_\lambda^i = 0$. For $p_i > 1$, the minimization reduces to solving

$$g_\lambda^i + \frac{\alpha_i}{2C^2} p_i \text{sign}(g_\lambda^i) |g_\lambda^i|^{p_i - 1} = C^{-2} \left[F_i K^* f + C^2 a^i - F_i K^* K^n a \right]_\lambda \,. \tag{3.9}$$

As it can be retraced in (Daubechies et al., 2004), the map $F_{\tau,p}(x) = x + \tau p \text{sign}(x) |x|^{p-1}$ is for any $p > 1$ a one-to-one map from $\mathbb{R}$ to itself, we thus find that for all $i = 1, \ldots, n$ and $\lambda \in \Lambda_i$,

$$g_\lambda^i = S_{\alpha_i/2C^2, p_i}(C^{-2} \left[F_i K^* f + C^2 a^i - F_i K^* K^n a \right]_\lambda) \,,$$

where $S_{\alpha_i/2C^2, p_i}$ is defined by $S_{\alpha_i/2C^2, p_i} := (F_{\alpha_i/2C^2, p_i})^{-1}$. For $p_i = 1$, let the sign function be set-valued (because of the non-differentiability of $|\cdot|$ at 0), i.e. $\text{sign}(t) = \pm 1$ for $t \gtrless 0$ and $\text{sign}(t) \in [-1, 1]$ for $t = 0$, leading to

$$g_\lambda^i + \frac{\alpha_i}{2C^2} \text{sign}(g_\lambda^i) \ni C^{-2} \left[F_i K^* f + C^2 a^i - F_i K^* K^n a \right]_\lambda \,. \tag{3.10}$$

23

In this case the associated operator $S_{\alpha_i/2C^2,1}$ is nothing than the well-known soft-shrinkage operator with threshold $\alpha_i/2C^2$. Introducing for some $h \in \ell_2$ the sequence-wise acting operator $S_{t,p_i}(h) = \{S_{t,p_i}(h_\lambda)\}_{\lambda \in \Lambda_i}$, we may define the following 'generalized' shrinkage operator for a vector of sequences $(g^1, \ldots, g^n) \in (\ell_2)^n$ and parameter vectors $t = (t_1, \ldots, t_n)$ and $p = (p_1, \ldots, p_n)$,

$$\mathbf{S}_{t,p}(g) = \left(S_{t_1,p_1}(g^1), \ldots, S_{t_n,p_n}(g^n) \right) .$$

With the latter shorthand notation the minimizer g of (3.6) can be written in the more clearly arranged form

$$g = \mathbf{S}_{\alpha/2C^2,p} \left(C^{-2} \left[(K^n)^* f + C^2 a - (K^n)^* K^n a \right] \right) . \tag{3.11}$$

The following proposition can be found in (Teschke, 2007) or can be retraced with the help of (Daubechies et al., 2004).

Proposition 3.1 *Suppose the operator K maps a Hilbert space $\mathcal{H}$ to another Hilbert space $\mathcal{H}'$ and is bounded by $\tilde{C}$. Furthermore, suppose we are given n frames where the respective frame operators F_i map $\mathcal{H}$ to ℓ_2 with upper frame bounds B_i. Assume, moreover, that f is an element of $\mathcal{H}'$ and $a \in (\ell_2)^n$. If $\Phi^{sur}(g;a)$ is defined as in (3.6) on $(\ell_2)^n$, then $\Phi^{sur}(g;a)$ has a unique minimizer in $(\ell_2)^n$. This minimizer is given by*

$$g = \mathbf{S}_{\frac{\alpha}{2C^2}} \left(C^{-2} \left[(K^n)^* f + C^2 a - (K^n)^* K^n a \right] \right) . \tag{3.12}$$

For all $h \in (\ell_2)^n$, one has

$$\Phi^{sur}(g + h; a) \geq \Phi^{sur}(g; a) + C^2 \|h\|^2_{(\ell_2)^n} .$$

This result directly carries over to iteration (3.7):

Corollary 3.1 *Make the same assumptions as in Proposition 3.1. Pick $g_0 \in (\ell_2)^n$ arbitrarily. Then the iterates of the algorithm (3.7) have the following explicit form*

$$g_{m+1} = \mathbf{S}_{\alpha/2C^2,p} \left(C^{-2} \left[K_A^* f + C^2 g_m - K_A^* K_A g_m \right] \right) . \tag{3.13}$$

As the final result of this section, it can be verified that the proposed iteration (3.7) converges in the norm of $(\ell_2)^n$.

Theorem 3.1 *Suppose the operator K maps a Hilbert space $\mathcal{H}$ to another Hilbert space $\mathcal{H}'$ and is bounded by $\tilde{C}$. Furthermore, suppose we are given n frames where the respective frame operators F_i map $\mathcal{H}$ to ℓ_2 with upper frame bounds B_i. Assume, moreover, that f is an element of $\mathcal{H}'$ and $a \in (\ell_2)^n$. Then the sequence of iterates*

$$g_{m+1} = \mathbf{S}_{\alpha/2C^2} \left(C^{-2} \left[(K^n)^* f + C^2 g_m - (K^n)^* K^n g_m \right] \right) , \quad m = 0, 1, 2, \ldots ,$$

with g_0 arbitrarily chosen in $(\ell_2)^n$, converges in norm to a minimizer of the functional

$$\Phi(g) = \|f - K^n g\|^2_{\mathcal{H}'} + \alpha \cdot |||g||| .$$

The complete proof of this theorem is quite lengthy and technical, for reader convenience we refer to (Daubechies et al., 2004) or (Teschke, 2007). Essentially the proof consists of two steps. At first, based on Opial's Theorem (see (Opial, 1967)), the weak convergence is shown. In a second step it is shown that the convergence holds also in norm.

3.3 Audio coding example

Within this section we show the usefulness of the proposed 'multi–frame' approach. We present two numerical experiments from different perspectives: convergence rates, sparsity achievement and approximation quality.

The overall configuration of our algorithm is as follows: for the sake of simplicity, we pick as our underlying frames a wavelet basis (Haar system) and a (non–local) Fourier basis only. Hence, $B_1 = B_2 = 1$. In the examples, we restrict ourselves to $K = I$. Consequently, the constant C in our Gaussian surrogate is not allowed to be equal or smaller than $\sqrt{2}$. We aim to achieve sparsity in both representations, i.e. we set $p_1 = p_2 = 1$. The variational problem is thus simply given by

$$\Phi(g^1, g^2) = \|f - (F_1^* g^1 + F_2^* g^2)\|^2 + \alpha_1 \|g^1\|_{\ell_1} + \alpha_2 \|g^2\|_{\ell_1} \, ,$$

and the minimization by Gaussian surrogates yields the following iteration

$$\begin{pmatrix} (g^1)_{m+1} \\ (g^2)_{m+1} \end{pmatrix} = \begin{pmatrix} S_{\alpha_1/2C^2,1} \left(C^{-2} \{ F_1 f + C^2 (g^1)_m - F_1 F_1^* (g^1)_m - F_1 F_2^* (g^2)_m \} \right) \\ S_{\alpha_2/2C^2,1} \left(C^{-2} \{ F_2 f + C^2 (g^2)_m - F_2 F_1^* (g^1)_m - F_2 F_2^* (g^2)_m \} \right) \end{pmatrix} .$$

Since we deal with bases only, the application of $F_1 F_1^*$, $F_1 F_2^*$, $F_2 F_1^*$, and $F_2 F_2^*$ simplifies the discrete decomposition and reconstruction schemes. If one really goes beyond bases, i.e. using frames, one indeed has to compute (approximate) all the (mixed) gram matrices. This might be of course costly but can be optimized by picking localized and reasonably incoherent frames. As an experimental observation, in case the frame generating analyzing atmos are not reasonably distinct, the scheme is not able to separate the signal components adequately, i.e. all the sequences g^i contain very similar informations.

A synthetic Example. In this example we have simulated a signal f that is a composition of two different components: a harmonic wave and noisy perturbation within the interval $[350, 400]$. As a sampled discrete vector it has a total number of 631 coefficients in the time–domain representation. This discrete vector is used as input for our algorithm. The results for $\alpha_1 = \alpha_2 = 0.2$ are visualized in Figure 8. We find that involving the Haar wavelet basis and the Fourier basis splits the signal in very sparse and well separated components. The sparseness evolution of the two individual components can be seen in the 'sparsity' plot in Figure 8 approving that the chosen frames meet quite nicely the signal structure.

Real data: "Glockenspiel". This data set represents a real audio signal consisting of tonal components and a sequence of (bell) attacks. We again try to apply Haar wavelet

and Fourier splitting. For $\alpha_1 = 0.02$ and $\alpha_2 = 0.01$ the results are shown in Figure 9. As expected, the Haar system captures all the bell attacks very well, and, moreover, the Fourier system the tonal components. The sparsity evolution graph shows the rapid decay of the number of wavelet coefficients which can be explained by a fast "bell attacks" localization process through the iteration.

We summarize, whenever the dictionary consists of complementary frames, the proposed algorithm produces a sparse representation in which the individual components overlap inconsiderably. However, a different choice of penalty weights would of course imply a different splitting of the signal: if $\alpha_1 >> \alpha_2$, then almost everything of the signal would be captured by the wavelet system and vice versa. The audio results can be downloaded from `http://www.zib.de/AG_InverseProblems/wav/`.
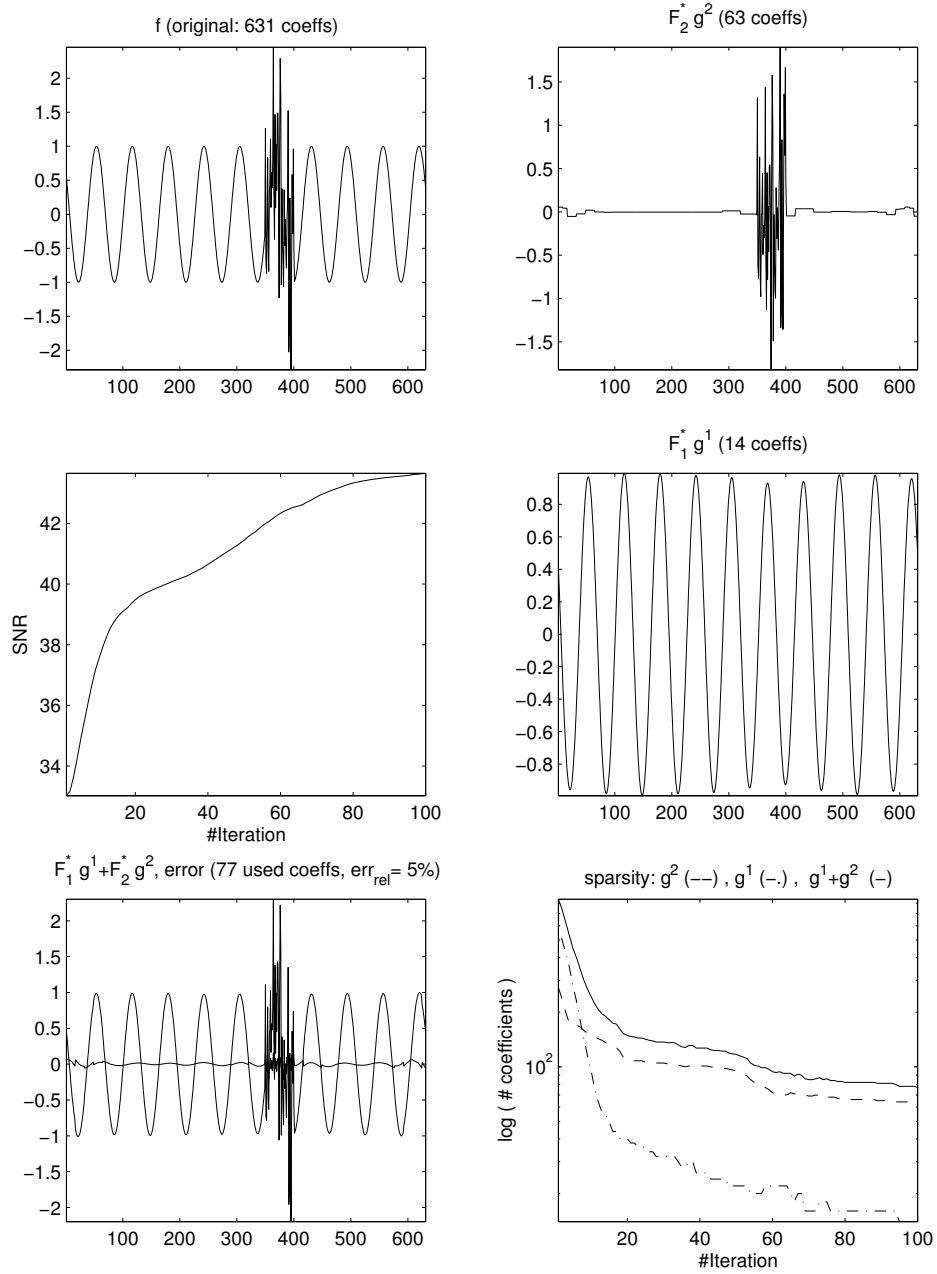
Figure 8: From top left to up right: synthetic data, Haar wavelet component (g^2) (in time domain) after 100 iterations, SNR evolution through the iteration process, Fourier component (g^1) (in time domain) after 100 iterations, reconstruction and error after 100 iterations, and sparsity evolution through the iteration process.
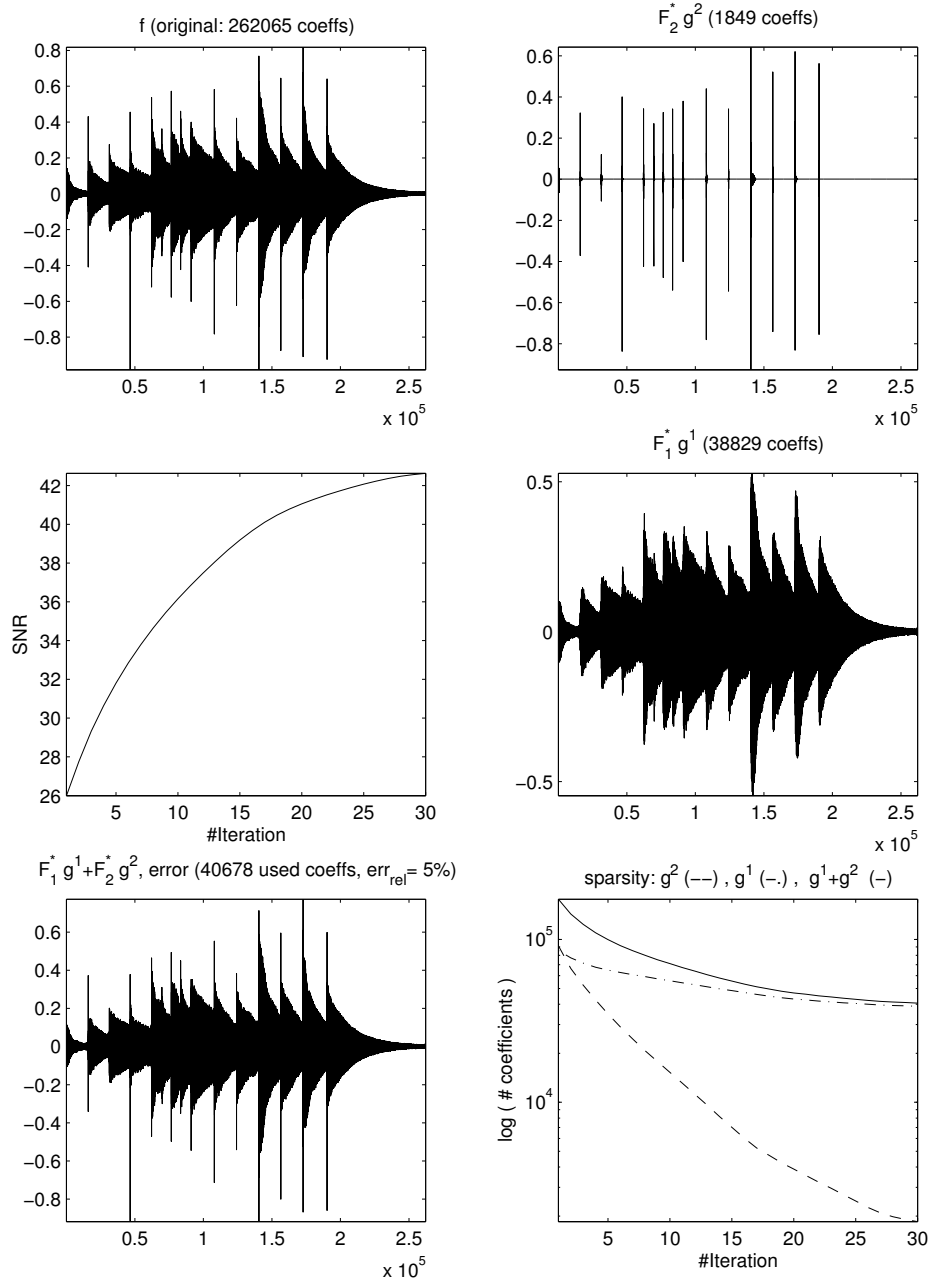
Figure 9: From top left to up right: "Glockenspiel" data, Haar wavelet component (g^2) (in time domain) after 30 iterations, SNR evolution through the iteration process, Fourier component (g^1) (in time domain) after 30 iterations, reconstruction after 30 iterations, and sparsity evolution through the iteration process.

4 Image restoration with general convex constraints

In Section 2 and Section 3 we have considered image restoration problems in which the constraints on the signal/image to be reconstructed could be directly expressed by means of basis or frame coefficients (or could be adequately replaced). However, for certain applications it might not be desirable to formulate the constraints in such a way. Often a more general description of the constraint is much better suited. We have seen in Section 2 (in the context of image decomposition) that in the setting of (Vese and Osher, 2003) and (Osher et al., 2003), the cartoon component of an image was modelled by a BV function. The BV penalty term was then replaced by a $B_1^1(L_1)$ term (amounting to a slightly stronger constraint) in order to write the problem in the elegant wavelet framework in which the proposed iteration scheme was easy to apply. It might now be interesting to see whether a similar iteration scheme can be executed when waiving the comfort of a wavelet framework and allowing the solution to be a BV function (or fulfilling some other general homogeneous convex constraint).

4.1 Preliminaries on general convex constraints

As before, we consider a functional of the form

$$\|f - Kv\|_{\mathcal{H}}^2 + 2\alpha J(v), \tag{4.1}$$

where $J(v) < \infty$, or even $J(v) < 1$ is the mathematical translation of the a priori knowledge (sometimes, we will use $\|\cdot\|$ for $\|\cdot\|_{\mathcal{H}}$). In what follows, we shall consider two different choices of $J(v)$, both adapted to the case where the inverse problem consists in deblurring and denoising a 2-dim. image, as in (Daubechies and Teschke, 2005), which was in turn, inspired by (Daubechies et al., 2004) and (Vese and Osher, 2003). Both approaches are natural sequels to (Daubechies and Teschke, 2005). In the first approach, we consider $J(v)$ of the same type as in (Daubechies and Teschke, 2005), but we put it in a more general framework, where $J(v)$ can be any positive, convex, one–homogeneous functional. An extensive discussion of such functionals, in much greater generality than what we present here, is given in (Combettes and Wajs, 2005). In order to be self contained, and to avoid introducing the full complexity of (Combettes and Wajs, 2005), we present here a sketch of a simpler version that suffices for our case (for a detailed discussion on the proof we refer the interested reader to (Daubechies et al., 2007)). In the second approach, $J(v)$ is the same as in (Rudin and Osher, 1994) and (Osher et al., 2003), but the numerical solution in (Osher et al., 2003) of a 4-th order nonlinear PDE is replaced by an iterative approach similar to (Daubechies et al., 2004) and (Daubechies and Teschke, 2005) (we also refer the reader to related prior work of (Bect et al., 2004)).

We assume that the functional to minimize takes the form (4.1), where J is a positive, convex and one–homogeneous functional. In this case, the variational problem can be recast as follows: Consider J^*, the Fenchel transform or so–called dual functional of J,

see (Ekeland and Témam, 1999; Rockafellar and Wets, 1998). Since J is positive and one–homogeneous, there exists a convex set C such that J^* is equal to the indicator function χ_C over C. In Hilbert space, we have total duality between convex sets and positive and one–homogeneous functionals, i.e. $J = (\chi_C)^*$, or

$$(\chi_C)^*(v) = \sup_{h \in C}\langle v, h \rangle = J(v) \ ;$$

see, e.g., (Ekeland and Témam, 1999; Aubert and Aujol, 2005; Chambolle, 2004; Combettes and Wajs, 2005). (Note: (Combettes and Wajs, 2005) gives a much more general and complete discussion; we restrict ourselves here to a simple situation, and only sketch the arguments. For a complete, detailed discussion, we refer the reader to (Combettes and Wajs, 2005).) We thus end up with the following reformulation of our problem: given some closed convex set $C \subset \mathcal{H}$ (on which we may still impose extra conditions, below), we wish to minimize

$$\mathcal{F}_C(v) = \|f - Kv\|_{\mathcal{H}}^2 + 2\alpha \sup_{h \in C}\langle v, h \rangle \ , \tag{4.2}$$

where we assume K to be a bounded operator from $\mathcal{H}$ to itself, with $\|K\| < 1$. We shall consider two particular cases in more detail.

Example 1. As in (Daubechies et al., 2004), a particular orthonormal basis $\{\phi_\lambda\}_{\lambda \in \Lambda}$ in $\mathcal{H}$ is preselected, and the prior is defined as

$$J(v) = \sum_{\lambda \in \Lambda}|\langle v, \phi_\lambda\rangle| \ .$$

This can, of course, be viewed as a special case of (4.2), since in this case

$$C = \{h \in L_{2,per}([0,1]^2); \ |\langle h, \phi_\lambda\rangle| \leq 1, \ \ \forall \lambda\}.$$

Similarly, the case with the prior

$$|v|_w = \sum_{\lambda \in \Lambda}w_\lambda|\langle f, \phi_\lambda\rangle| \ , \ \text{ with } \inf_{\lambda \in \Lambda}w_\lambda > 0 \ ,$$

fits also into the framework of (4.2), with C now defined by

$$C = \{h \in L_{2,per}([0,1]^2); |\langle h, \phi_\lambda\rangle| \leq w_\lambda^{-1}, \ \ \forall \lambda\} \ .$$

When $K \neq I$ and the problem is ill-posed, the resulting minimization scheme amounts to Landweber iteration with thresholding applied in each step.

Example 2. In the BV regularization framework, (Rudin et al., 1992), (Rudin and Osher, 1994), one considers functionals of the form

$$\|f - Kv\|_{L_2(\Omega)}^2 + 2\alpha \int_\Omega |\nabla v| \tag{4.3}$$

and minimizes over all possible $v \in L_2(\Omega)$. Expressing this functional by means of a convex set C one discovers that C is the L_2-closure of

$$C = \{h \in L_{2,per}([0,1]^2); \ h = \operatorname{div} g, \text{ where } g \text{ is a 2d field in } \mathrm{C}_c^1(\Omega)^2 \text{ that satisfies}$$

$$\|g\|_{\ell_\infty} = \sup_{(x,y) \in [0,1]^2} (|g_1(x,y)|^2 + |g_2(x,y)|^2)^{1/2} \le 1\} \ ,$$

i.e. we may again write

$$\sup_{h \in C} \langle v, h \rangle = \int_\Omega |\nabla v| = |v|_{BV} \ ;$$

for details on the structure of C we refer the reader to (Evans and Gariepy, 1991; Ambrosio et al., 2000). It turns out that results on iterative strategies developed for Example 1 carry over to the BV case and that much of the analysis elaborated in (Daubechies et al., 2004) can be generalized to the minimization of (4.3).

4.1.1 Reformulation of the problem

We shall assume that C is a closed convex set in $\mathcal{H}$, C is symmetric, i.e. $h \in C \Rightarrow -h \in C$, and there exists finitely many vectors $a_1, \ldots a_N \in \mathcal{H}$, and $r > 0$ so that $B_r(0) \cap \{a_1, \ldots a_N\}^\perp \subset C$ (i.e. $\{h : \ \langle h, a_i \rangle = 0 \ ; i = 1, \ldots, N \text{ and } \|h\| < r\} \subset C$). Note: we introduce the finite–dimensional subspace to which C is orthogonal for two reasons. First, there are cases of interest in which C consists of functions that have zero mean in $[0,1]^2$, e.g. if C contains only divergences of smooth periodic fields. Second, it will make it easier to restrict ourselves to only fine scale functions, below.

Defining the functionals

$$L(v, h) := \|f - Kv\|_{\mathcal{H}}^2 + 2\alpha \langle v, h \rangle,$$

we can rewrite $\inf_{v \in \mathcal{H}} \mathcal{F}_C(v)$ as

$$\inf_{v \in \mathcal{H}} \sup_{h \in C} L(v, h) \ . \tag{4.4}$$

Note that $L(v, h)$ is continuous in both arguments, it is also convex with respect to v, concave with respect to h. This means that (provided some technical conditions are fulfilled, see (Boyd and Vandenberghe, 2005), (Hiriart-Urruty and Lemaréchal, 1993), (Ekeland and Témam, 1999) or (Daubechies et al., 2007)) we can apply the minimax theorem, which allows us to interchange inf and sup in (4.4). In this case the minimax theorem moreover asserts that inf and sup are achieved, i.e. the inf is a min, the sup is a max.

4.2 Solving the inverse problem for convex penalization

Although the case where K^*K does not have a bounded inverse, i.e. where the inverse problem is ill–posed is of most interest to us, we start by sketching the approach in the

easier well–posed case.

Theorem 4.1 *Suppose that all assumptions made above hold true, and K^*K has bounded inverse in its range. If we define $A := (K^*K)^{-1/2}$ and, for an arbitrary closed convex set $\mathcal{K} \subset \mathcal{H}$, $S_{\mathcal{K}} := Id - P_{\mathcal{K}}$, where $P_{\mathcal{K}}$ is the (nonlinear) projection on $\mathcal{K}$, i.e. $P_{\mathcal{K}}\varphi = \arg\min_{h \in C} \|h - \varphi\|$, then the minimizing v is given by*

$$v = AS_{\alpha AC}\{AK^*g\} \ .$$

An obvious example is the case where we just need to denoise an image, without deblurring:

Example 3. Consider the denoising problem with an ℓ_1–constraint in the basis $\{\phi_\lambda\}_{\lambda \in \Lambda}$. In this case $K = Id$, so that $A = Id$ as well, and

$$C = \{v; \ \sup_{\lambda} |\langle v, \phi_\lambda \rangle| \leq 1\} \ .$$

Moreover, in the real case we have

$$\langle P_C v, \phi_\lambda \rangle = \begin{cases} \langle v, \phi_\lambda \rangle & \text{if } |\langle v, \phi_\lambda \rangle| \leq 1 \\ \text{sign}\langle v, \phi_\lambda \rangle & \text{if } |\langle v, \phi_\lambda \rangle| > 1 \ . \end{cases}$$

This implies that $S_{\alpha AC} \circ AK^*$ is exactly the soft thresholding operator

$$\langle S_{\alpha AC}(AK^*f), \phi_\lambda \rangle = \mathsf{S}_\alpha(\langle f, \phi_\lambda \rangle) \ .$$

In the complex case, we have

$$\langle P_C v, \phi_\lambda \rangle = \begin{cases} \langle v, \phi_\lambda \rangle & \text{if } |\langle v, \phi_\lambda \rangle| \leq 1 \\ \frac{\langle v, \phi_\lambda \rangle}{|\langle v, \phi_\lambda \rangle|} & \text{if } |\langle v, \phi_\lambda \rangle| > 1 \ , \end{cases}$$

and the $S_{\alpha AC} \circ AK^*$ reduces to the "complex soft thresholding operator", i.e.

$$\langle S_{\alpha AC}(AT^*f), \phi_\lambda \rangle = \mathsf{S}_\alpha^c(\langle f, \phi_\lambda \rangle) = \begin{cases} \langle f, \phi_\lambda \rangle - \alpha\frac{\langle f, \phi_\lambda \rangle}{|\langle f, \phi_\lambda \rangle|} & \text{if } |\langle f, \phi_\lambda \rangle| > \alpha \\ 0 & \text{if } |\langle f, \phi_\lambda \rangle| \leq \alpha \ . \end{cases}$$

In the most interesting problems, the operator K^*K does not have a bounded inverse. We can then use the surrogate functionals introduced in (Daubechies et al., 2004). We replace (4.1) by a family of surrogate functionals

$$\begin{aligned} \mathcal{G}_{n,C}(v) &= \mathcal{F}_C(v) + \|v_n - v\|^2 - \|K(v_n - v)\|_{\mathcal{H}}^2 \\ &= -2\langle v, K^*f \rangle + 2\alpha \sup_{h \in C}\langle v, h \rangle - 2\langle v, v_n \rangle + 2\langle v, K^*Kv_n \rangle \\ &\quad + \|v\|^2 + \|f\|^2 + \|v_n\|^2 - \|Kv_n\|_{\mathcal{H}}^2 \ , \end{aligned}$$

and we have

Proposition 4.1 *Let C be as assumed in Section 4.1.1. Then the minimizer of $\mathcal{G}_{n,C}$ is given by*

$$v_{n+1} := (Id - P_{\alpha C})(v_n + K^* f - K^* K v_n) \ . \qquad (4.5)$$

We mentioned above that (Combettes and Wajs, 2005) contains an extensive discussion, including (not easily verifiable) conditions that ensure strong convergence for an iteration of type (4.5). However, the full generality of (Combettes and Wajs, 2005) makes it less easy to read if one is mainly interested in the special case discussed here. Since the iteration is very similar to the one in (Daubechies et al., 2004), a very similar strategy for the proof of convergence holds as well. It can be retraced in (Daubechies et al., 2007) that up to strong convergence the techniques apply in almost the same way (weak convergence is achieved by applying Opial's Theorem, see (Opial, 1967)). In order to achieve norm convergence, we have to pay more attention to the structure of C, however.

One can argue that weak convergence suffices for practical purposes, because every numerical computation is always finite–dimensional so that weak and strong (i.e. norm) convergence of the v_n are equivalent. However, it is often useful to establish norm convergence for the infinite dimensional Hilbert space as well, since this then implies that the rate of convergence, and the other constants involved, do not "blow up" as the dimensionality of the discretization increases.

To obtain norm convergence, we need to do some more work. It can be verified in (Daubechies et al., 2007), that the have following facts :

- $v_n \xrightarrow{\text{weak}} \bar{v}, \quad$ for $n \to \infty$,

- $\bar{v} = \bar{v} + K^* f - K^* K \bar{v} - P_{\alpha C}(\bar{v} + K^* f - K^* K \bar{v})$,

- $v_{n+1} = v_n + K^* f - K^* K v_n - P_{\alpha C}(v_n + K^* f - K^* K v_n)$,

- $\|v_{n+1} - v_n\| \to 0, \quad$ for $n \to \infty$.

Defining

$$u_n := v_n - \bar{v} \ \text{ and } \ w := \bar{v} + K^* f - K^* K \bar{v}, \qquad (4.6)$$

we can recast the facts as follows:

$$u_n \xrightarrow{\text{weak}} 0, \quad \text{for } n \to \infty$$

$$\|P_{\alpha C}(v) - P_{\alpha C}(w + u_n - K^* K u_n) - K^* K u_n\| \to 0, \quad \text{for } n \to \infty \ .$$

We can then apply, without any change, Lemmas 3.15, 3.17 of (Daubechies et al., 2004), leading to

$$\|K^* K u_n\| \to 0, \quad \text{for } n \to \infty,$$

33

so that we obtain the equivalent formulation

$$u_n \xrightarrow{\text{weak}} 0, \quad \text{for } n \to \infty$$

$$\|P_{\alpha C}(w) - P_{\alpha C}(w + u_n)\| \to 0, \quad \text{for } n \to \infty.$$

(4.7)

To obtain norm convergence of the v_n, we must establish $\|u_n\| \to 0$. For general convex sets C the conditions (4.7), where $\alpha > 0$ and $w \in \mathcal{H}$ are arbitrary (but fixed) actually do not imply norm convergence of the u_n to 0. Abstract sufficient and necessary conditions for norm convergence are given in (Combettes and Wajs, 2005); the following theorem (for a proof see (Daubechies et al., 2007)) gives a more concrete restriction on C under which we can establish norm convergence.

Theorem 4.2 *Suppose* $u_n \xrightarrow{\text{weak}} 0$ *and* $\|P_{\alpha C}(w) - P_{\alpha C}(w + u_n)\| \to 0$. *Moreover, assume that* u_n *is orthogonal to* w, $P_C(w)$. *If for some sequence* γ_n *(with* $\gamma_n \to \infty$*) the convex set* C *satisfies* $\gamma_n u_n \in C$ *then* $\|u_n\| \to 0$.

Unfortunately, this theorem is not sufficiently strong to be applied to the BV–functional of *Example 2*, above. Without going in full detail, we sketch here how it (just) falls short.

The set C in *Example 2* is (loosely speaking) the set of all divergences of 2-dim. fields that are uniformly bounded by 1. It contains, in particular, the functions

$$\begin{aligned}
h_\mathbf{n}(x, y) \quad &:= \quad \sqrt{2}\pi(|n_1| + |n_2|)e^{2\pi i(n_1 x + n_2 y)} \\
&= \quad -i\mathrm{div}\left(\frac{1}{\sqrt{2}}\mathrm{sign}(n_1)e^{2\pi i(n_1 x + n_2 y)}, \frac{1}{\sqrt{2}}\mathrm{sign}(n_2)e^{2\pi i(n_1 x + n_2 y)}\right),
\end{aligned}$$

where $|n_1| + |n_2| \neq 0$. Because C is closed and convex, it also contains all the

$$\sum_{\mathbf{n} \in \mathbb{Z}^2} \alpha_\mathbf{n} h_\mathbf{n} \ ,$$

with $\sum_{\mathbf{n} \in \mathbb{Z}^2} \alpha_\mathbf{n} = 1$. Suppose now (just for the sake of simplifying the argument, which can also be made, a bit more lengthily, without this assumption) that

$$\|P_C(u_n)\| \to 0 \text{ as } n \to \infty \ ,$$

i.e. that the condition

$$\|P_{\alpha C}(w) - P_{\alpha C}(w + u_n)\| \to 0 \text{ as } n \to \infty$$

holds true for $w = 0$. That would mean that, for all $g \in C$

$$\lim_{n \to \infty} \langle u_n - P_C(u_n), g - P_C(u_n) \rangle$$

$$= \quad \lim_{n \to \infty} \langle u_n, g \rangle + \|P_C(u_n)\|^2 - \langle u_n, P_C(u_n) \rangle - \langle P_C(u_n), g \rangle$$

$$= \quad \lim_{n \to \infty} \langle u_n, g \rangle \ ,$$

34

which implies that $\lim_{n\to\infty}\langle u_n, g\rangle$ is nonpositive. Since the same is true for $-g \in C$, if follows that $\lim_{n\to\infty}\langle u_n, g\rangle = 0$ for all $g \in C$. Consequently, $\langle u_n, h_{\mathbf{k}}\rangle \to 0$ as $n \to \infty$, or even, for all sequences $(\alpha_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^2}$ with $\sum_{\mathbf{k}\in\mathbb{Z}^2}\alpha_{\mathbf{k}} = 1$, $\alpha_{\mathbf{k}} \in [0,1]\ \forall \mathbf{k}$,

$$\sum_{\mathbf{k}} \alpha_{\mathbf{k}}(|k_1| + |k_2|)\langle u_n, e_{\mathbf{k}}\rangle \to 0 \text{ as } n \to \infty \ ,$$

where $e_{\mathbf{k}}(x, y) = e^{2\pi i(k_1 x + k_2 y)}$. This just misses ensuring that

$$\sum_{\mathbf{k}} |\langle u_n, e_{\mathbf{k}}\rangle|^2 \to 0 \text{ as } n \to \infty \ .$$

This concludes our theoretical analysis of our first case described in the introduction, i.e. the case where $J(f)$ in (4.1) is convex.

4.3   Numerical Illustrations

4.3.1   Iterative algorithm for PDE–based deblurring and denoising

In the framework of (Rudin and Osher, 1994), the edge-preserving energy functional is of the form

$$\mathcal{F}(v) = \|f - Kv\|_{L_2(\Omega)}^2 + 2\alpha \int_\Omega \phi(\nabla f)dx \ ; \tag{4.8}$$

where the potential $\phi : \mathbb{R}^2 \to \mathbb{R}$ is typically a positive continuous function, with at most linear growth at infinity. Convex examples include (note that, only for illustration reasons, we also give examples beyond the one–homogeneous case)

- $\phi(\xi) = |\xi|$ (the total variation minimization (Rudin et al., 1992), (Rudin and Osher, 1994)),

- $\phi(\xi) = |\xi_1| + |\xi_2|$,

- $\phi(\xi) = \sqrt{1 + |\xi|^2}$ (the function of minimal surfaces (Aubert and Vese, 1997), (Vese, 2001)),

- $\phi(\xi) = \log\cosh(1 + |\xi|^2)$, or

- $\phi(\xi) = \begin{cases} \frac{1}{2}|\xi|^2 \text{ if } |\xi| \leq 1 \\ |\xi| - \frac{1}{2} \text{ if } |\xi| \geq 1 \end{cases}$ (used in (Demengel and Temam, 1984), (Chambolle and Lions, 1997)).

In the non-convex case, examples of the potential ϕ are

- $\phi(\xi) = \frac{|\xi|^p}{1+|\xi|^p}$ or

$$\frac{1}{60}$$

0	0	0	0	0	0.25	0	0	0	0	0
0	0	0.25	0.5	0.5	0.5	0.5	0.5	0.25	0	0
0	0.25	0.5	1	1	1	1	1	0.5	0.25	0
0	0.5	1	1	1	1	1	1	1	0.5	0
0	0.5	1	1	1	1	1	1	1	0.5	0
0.25	0.5	1	1	1	1	1	1	1	0.5	0.25
0	0.5	1	1	1	1	1	1	1	0.5	0
0	0.5	1	1	1	1	1	1	1	0.5	0
0	0.25	0.5	1	1	1	1	1	0.5	0.25	0
0	0	0.25	0.5	0.5	0.5	0.5	0.5	0.25	0	0
0	0	0	0	0	0.25	0	0	0	0	0

Table 3: Spatial discretization of the blur operator K.

- $\phi(\xi) = \log(1 + |\xi|^p)$, with $p = 1$ or $p = 2$ for instance, see (Geman and Geman, 1984), (Geman and Reynolds, 1992), Perona-Malik (Perona and Malik, 1990), and more recently (Aubert and Vese, 1997).

Let us now restrict again to the one–homogeneous case and assume in addition that ϕ is differentiable. Then the Euler-Lagrange equation associated with the minimization problem (4.8), that must be satisfied by a minimizer v, if such a minimizer exists, is given by

$$K^*Kv - K^*f = \alpha \text{div}\Big(\nabla_\xi \phi(\nabla v)\Big), \text{ in } \Omega, \tag{4.9}$$

where $\nabla\phi_\xi = (\phi_{\xi_1}, \phi_{\xi_2})$, and with the boundary conditions $\nabla_\xi \phi(\nabla v) \cdot \vec{n} = 0$ on $\partial\Omega$, where $\vec{n}$ is the unit exterior normal to the boundary. In the case $\alpha > 0$, the partial differential equation (4.9) is non-linear for the examples of potential ϕ given above. Moreover, the presence of the term K^*Kf makes it computationally expensive and numerically nontrivial.

In order to overcome these problems, we propose here to not directly solve (4.9) numerically, but to apply the surrogate functional algorithm (see (Daubechies et al., 2004), or the previous sections), i.e. we construct a sequence of iterates v_n that approximate v, without having to invert K^*K at every iteration. On the other hand, the direct implementation of the projection $P_{\alpha C}$ associated to our minimization is rather complicated; in this case we prefer to avoid it by switching to an expression based on the Euler-Lagrange equation. The total iteration goes thus as follows: start with an initial v_0; find v_n, $n > 0$ as a minimizer of the surrogate functionals

$$\mathcal{G}_{n-1}(v_n) = \|f - Kv_n\|^2_{L_2(\Omega)} - \|Kv_n - Kv_{n-1}\|^2_{L_2(\Omega)} + \|v_n - v_{n-1}\|^2_{L_2(\Omega)} + 2\alpha \int_\Omega \phi(\nabla v_n), \tag{4.10}$$

where we have assumed that $\|K^*K\| < 1$. The associated Euler-Lagrange equation in v_n, now easily solved in practice, is:

$$v_n = v_{n-1} + K^*v - K^*Kv_{n-1} + \alpha \text{div}\Big(\nabla_\xi \phi(\nabla v_n)\Big), \tag{4.11}$$

Figure 10: Top left: original image. Top right: blurred version.

together with the same boundary conditions. One then simply carries out this iterative algorithm to find (an approximation to) desired minimizer.

4.3.2 Comparison of the iteration schemes with Besov and BV constraints

In order to illustrate the capabilities and differences with respect to reconstruction quality and computational cost, we present some numerical results of the Besov (wavelet framework) and the BV approach. We assume that the linear degradation model $f = Kv + e$, where f is the given data, as a square integrable function in $L_2(\Omega)$, v is the unknown true image, e is additive noise of zero mean. The operator $K : L_2(\Omega) \to L_2(\Omega)$ models a linear and continuous degradation operator, by a convolution with a Gaussian kernel.

In the first approach, we have chosen a wavelet frame that is simply given by a translation invariant wavelet system and have applied the iterative deconvolution scheme of Section 2 (see also (Daubechies et al., 2004; Daubechies and Teschke, 2004, 2005)). As the example image we consider a fingerprint and its blurred version, see Figure 10. The results obtained with iteration from the previous section are visualized in Figure 11 and the convergence rates are given in Table 4. The blur operator T used in the experiments has the discrete spatial representation given in Table 3.

The blur convolution is easily implemented as a multiplication in Fourier domain, which means that we switch between the wavelet and Fourier representation at every step of the iteration process.

Next, we present numerical results for the second (PDE) approach. In Figure 12 we show the results of the iterative algorithm (4.11) on the same blurred and noisy image. For comparison with the purely PDE–based method (without the iterative approach corresponding to surrogate functionals) we show in Figure 13 the end results of methods (4.11) and (4.9); they look very similar. Table 5 lists the CPU time and the relative RMSE for the first 5000 iterations of both methods, illustrating that the surrogate functional method produces a better error decay for the same amount of CPU time. (These
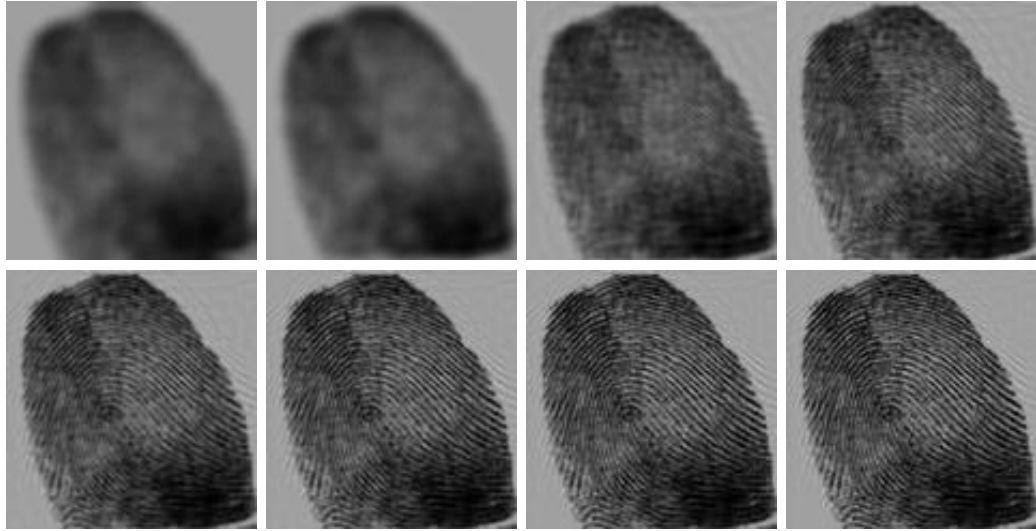
Figure 11: Top left to bottom right: blurred image, several iterates using the wavelets scheme: 1st, 100th, 500th, 1000th, 2000th, 3000th, 4000th.

two computations were carried out on the same machine; note that the numerical results in Table 11 were obtained on a different computer and should thus not be compared with this Table.)

Figure 12: Top left to bottom right: blurred image, several iterates using the surrogate functional iteration (4.11) with $\phi(\xi) = \sqrt{\epsilon + |\xi|^2}$ (total variation minimization with regularization): 1st, 100th, 500th, 1000th, 2000th, 3000th, 4000th.

Figure 13: Deblurring results obtained using the models (4.11) left and (4.9) right, with $\phi(\xi) = \sqrt{\epsilon + |\xi|^2}$ (total variation minimization with regularization).

ITER	CPU	RMSE $\|v_n - v_{orig}\|/\|v_{orig}\|$
1	0.28600841	0.17322202591876
100	11.0832407	0.16619377802673
200	21.3800888	0.16053661204032
350	36.9959693	0.15341521024097
500	52.2572860	0.14754897184399
1000	104.031806	0.13476786059559
2000	207.406313	0.12458884053496
3000	312.880765	0.11962522471006
4000	419.051499	0.11663033283685
5000	524.362921	0.11388107705039

Table 4: Convergence rates for the wavelet frame–based shrinkage algorithm. We give the number of iterations, the CPU time and the corresponding relative RMSE, applied to the blurry fingerprint image.

With surrogate functionals

ITER	CPU	RMSE$\|v_n - v_{orig}\|/\|v_{orig}\|$
1	0	0.256365806
100	53.7299995	0.22643654
200	107.440002	0.213100523
350	188.470001	0.19897379
500	268.699982	0.188747868
1000	542.880005	0.168787047
2000	1085.82996	0.153525516
3000	1625.08997	0.148305818
4000	2151.29004	0.146324202
5000	2684.62988	0.145639017

Without surrogate functionals

ITER	CPU	RMSE$\|v_n - v_{orig}\|/\|v_{orig}\|$
1	0.400000036	0.257728785
100	55.1000023	0.236117765
200	109.139999	0.226667866
350	191.959991	0.216331303
500	276.119995	0.208168939
1000	551.720032	0.189164072
2000	1105.32996	0.169159457
3000	1666.43994	0.159271181
4000	2234.70996	0.15375042
5000	2790.92993	0.150457606

Table 5: Comparison of the convergence rates for PDE–based approach: for both algorithms (the classical PDE iteration and the Gaussian surrogate functional iteration), we give the number of iterations, the CPU time and the corresponding relative RMSE, applied to the blurry fingerprint image, using the total variation minimization. We notice that the new method using the surrogate functionals converges faster to the restored image: the relative RMSE $\|v_n - v_{orig}\|/\|v_{orig}\|$ hits the value 0.15 at 2.500 iterations instead of 5.000, and uses a CPU time of ~ 1300 instead of 2790; there seems thus to be a speed–up factor 2.

5 Hybrid wavelet–PDE image restoration schemes

In Section 2, we have constructed a wavelet–based scheme that solves the variational problem (2.3),

$$\inf_{u,v} \mathcal{F}_f(v,u) \ , \quad \text{where}$$

$$\mathcal{F}_f(v,u) = \|f - K(u+v)\|^2_{L_2(\Omega)} + \gamma \|v\|^2_{H^{-1}(\Omega)} + 2\alpha |u|_{B^1_1(L_1(\Omega))} \ ,$$

in a numerically very efficient way. As discussed in Section 2.4, non–redundant wavelet filtering often creates artifacts that manifest themselves as ringing and edge blurring. The suggested way to reduce these artifacts while keeping the computational cost at some very low level was given by introducing redundant wavelet systems. Another way of keeping sharp edges, as already mentioned in Section 2.1, is to model the 'cartoon' part of the image as a function of bounded variation. Up to the linear operator K, this coincides with (2.2),

$$\inf_{(v,u)} \mathcal{E}_f(v,u) = \|f - K(u+v)\|^2_{L_2(\Omega)} + \gamma \|v\|^2_{H^{-1}(\Omega)} + 2\alpha |u|_{BV(\Omega)} \tag{5.1}$$

and was under consideration in (Vese and Osher, 2003; Osher et al., 2003). But instead of solving (5.1) in a pure PDE fashion (as done in (Vese and Osher, 2003; Osher et al., 2003)), we propose a combined wavelet–PDE scheme, that keeps the advantage of wavelets to well represent oscillatory patterns and simple minimization, with the advantage of non–linear PDE formulation, that keeps sharp edges and representation of functions of bounded variation.

In the proposed alternating scheme, the minimization in u will be solved in a PDE function, by finite differences, while the minimization in v will be solved in a wavelets function. The data function f is known in the spatial domain $f(x_1, x_2) \approx f_{i,j}$, but also in the frequency domain, by its wavelets coefficients $(f_\lambda)_{\lambda \in J}$. Let us assume $K = I$ for the moment. Keeping $v = v^n$ fixed, $n \geq 0$, we compute $u = u^n$ minimizing

$$\int_\Omega |(f-v) - u|^2 dx + 2\alpha \int_\Omega |\nabla u| dx,$$

solving the Euler-Lagrange equation by finite differences

$$u = (f-v) + \alpha \text{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

Now keep $u = u^n$ fixed, compute its wavelet coefficients $(u_\lambda)_{\lambda \in J}$, and compute $v = v^{n+1}$ minimizing with respect to v

$$\sum_{\lambda \in J} \left(|(f_\lambda - u_\lambda) - v_\lambda|^2 + \gamma 2^{-2|\lambda|} |v_\lambda|^2 \right).$$

This leads to decoupled one-dimensional minimizations, and gives the desired $v^{n+1} = \tilde{v}_{\gamma,\alpha}$ as before, from its wavelet coefficients v_λ. Then the steps are repeated.

5.1 Combined wavelet-PDE scheme in the presence of blur

Consider now the problem

$$\inf_{(v,u)} \mathcal{G}_f(v,u) = \|f - K(u+v)\|_{L_2(\Omega)}^2 + \gamma\|v\|_{H^{-1}(\Omega)}^2 + 2\alpha|u|_{BV(\Omega)}, \qquad (5.2)$$

in the combined wavelet-PDE scheme, and we would like to apply again the surrogate functionals for the minimization.

One way to minimize this, and avoiding to invert operators involving K^*K at every iteration, is to consider the unknown pair (v,u) and to directly apply the surrogate functionals approximation: knowing (v_{n-1}, u_{n-1}), find (v_n, u_n) minimizer of

$$\mathcal{G}_{f,n}(v_n, u_n) = \|f - K(u_n + v_n)\|_{L_2(\Omega)}^2 + \gamma\|v_n\|_{H^{-1}(\Omega)}^2 + 2\alpha|u_n|_{BV(\Omega)}$$
$$-\|K(u_n + v_n) - K(u_{n-1} + v_{n-1})\|_{L_2(\Omega)}^2 + \mu\|(u_n + v_n) - (u_{n-1} + v_{n-1})\|_{L_2(\Omega)}^2$$

where μ is such that $\|K^*K\| \leq \mu$.

5.1.1 Characterization of minimizers

Theoretically, we consider slightly modified $\mathcal{F}_f(v,u)$ and $\mathcal{E}_f(v,u)$, as

$$\mathcal{G}_f(v,u) = \|f - K(u+v)\|_{L_2(\Omega)}^2 + \gamma\|v\|_{H^{-1}(\Omega)} + 2\alpha|u|_{B(\Omega)}, \qquad (5.3)$$

where $|u|_{B(\Omega)}$ is one of the semi-norms $|u|_{BV(\Omega)}$ or $|u|_{B_1^1(L_1(\Omega))}$. In this slightly modified case, we have the following characterization of minimizers (inspired from (Meyer, 2002; Le and Vese, 2005)): let $f \in L_2(\Omega)$, $u \in B(\Omega) \subset L_2(\Omega)$, $v \in L_2(\Omega) \cap H^{-1}(\Omega)$. We introduce the texture norm

Definition 5.1 *Given* $w \in L_2(\Omega)$, γ, $\alpha > 0$, *define*

$$\|w\|_* = \sup_{g \in B(\Omega), h \in H^{-1}(\Omega) \cap L_2(\Omega)} \left\{ \frac{(w, g+h)}{2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)}}, \ 2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)} \neq 0 \right\},$$

where $(\cdot, \cdot)$ *is the* $L_2(\Omega)$ *inner product.*

Then, we can show the following characterization of minimizers for the two models:

Theorem 5.1 *Let* $w = f - K(u+v)$. *Then*

1. $\|K^*f\|_* \leq \frac{1}{2}$ *if and only if* $(v,u) = (0,0)$ *is a minimizer of* (5.3).

2. *Suppose* $\|K^*f\|_* > \frac{1}{2}$. *Then* (v, u) *is a minimizer of* (5.3) *if and only if*

$$\|K^*w\|_* = \frac{1}{2} \ \text{and} \ (K^*w, u + v) = \frac{1}{2}(2\alpha|u|_{B(\Omega)} + \gamma\|v\|_{H^{-1}(\Omega)}).\tag{5.4}$$

Proof. If the model (5.3) yields minimizers $u = 0$ and $v = 0$, then for any $g \in B(\Omega)$, $h \in H^{-1}(\Omega) \cap L_2(\Omega)$,

$$\|f\|_{L_2(\Omega)}^2 \le 2\alpha|g|_{B(\Omega)} + \|f - K(g + h)\|_{L_2(\Omega)}^2 + \gamma\|h\|_{H^{-1}(\Omega)}.\tag{5.5}$$

Substituting in (5.5) g by ϵg and h by ϵh, and taking $\epsilon \to 0^+$, we obtain

$$|(K^*f, g + h)| \le \frac{1}{2}(2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)}).\tag{5.6}$$

By the definition of $\|.\|_*$, therefore $\|K^*f\|_* \le \frac{1}{2}$.

For the converse property in 1., assume that $\|K^*f\|_* \le \frac{1}{2}$. Then, for any $g \in B(\Omega)$ and $h \in H^{-1}(\Omega) \cap L_2(\Omega)$, with $2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)} \ne 0$, we have

$$(K^*f, g + h) \le (2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)})\|f\|_* \le \frac{1}{2}(2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)}).$$

We also have

$$2\alpha|g|_{B(\Omega)} + \|f - K(g + h)\|_{L_2(\Omega)}^2 + \gamma\|h\|_{H^{-1}(\Omega)}$$
$$= \ 2\alpha|g|_{B(\Omega)} + \|f\|_{L_2(\Omega)}^2 - 2(K^*f, g + h) + \|K(g + h)\|_{L_2(\Omega)}^2 + \gamma\|h\|_{H^{-1}(\Omega)}$$
$$\ge \ 2\alpha|g|_{B(\Omega)} + \|f\|_{L_2(\Omega)}^2 - (2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)}) + \|K(g + h)\|_{L_2(\Omega)}^2 + \gamma\|h\|_{H^{-1}(\Omega)}$$
$$= \ \|f\|_{L_2(\Omega)}^2 + \|K(g + h)\|_{L_2(\Omega)}^2 \ge \|f\|_{L_2(\Omega)}^2 = \mathcal{G}_f(0, 0).$$

Therefore, $u = 0$ and $v = 0$ gives the optimal decomposition in this case.

Now suppose $\|K^*f\|_* > \frac{1}{2}$. Let (v, u) be an optimal decomposition given by (5.3). We have $u \not\equiv 0$ or $v \not\equiv 0$. For $g \in B(\Omega)$, $h \in H^{-1}(\Omega) \cap L_2(\Omega)$, and $\epsilon \in \mathbb{R}$,

$$\mathcal{G}_f(v + \epsilon h, u + \epsilon g) \ = \ 2\alpha|u + \epsilon g|_{B(\Omega)} + \|w - \epsilon K(g + h)\|_{L_2(\Omega)}^2 + \gamma\|v + \epsilon h\|_{H^{-1}(\Omega)}$$
$$\ge \ 2\alpha|u|_{B(\Omega)} + \|w\|_{L_2(\Omega)}^2 + \gamma\|v\|_{H^{-1}(\Omega)},\tag{5.7}$$

and thus, by triangle inequality,

$$2\alpha|u|_{B(\Omega)} + \|w\|_{L_2(\Omega)}^2 + \gamma\|v\|_{H^{-1}(\Omega)}$$
$$\le \ 2\alpha|u|_{B(\Omega)} + 2\alpha|\epsilon||g|_{B(\Omega)} + \|w - \epsilon K(g + h)\|_{L_2(\Omega)}^2 + \gamma\left(\|v\|_{H^{-1}(\Omega)} + |\epsilon|\|h\|_{H^{-1}(\Omega)}\right) \ .$$

Therefore,

$$\|w\|_{L_2(\Omega)}^2 \le 2\alpha|\epsilon||g|_{B(\Omega)} + \|w - \epsilon K(g + h)\|_{L_2(\Omega)}^2 + \gamma|\epsilon|\|h\|_{H^{-1}(\Omega)},$$

and thus,

$$\|w\|^2_{L_2(\Omega)} \leq 2\alpha|\epsilon||g|_{B(\Omega)} + \left(\|w\|^2_{L_2(\Omega)} - 2\epsilon(K^*w, g+h) + \epsilon^2\|K(g+h)\|^2_{L_2(\Omega)}\right) + \gamma|\epsilon|\|h\|_{H^{-1}(\Omega)},$$

yielding

$$2\alpha|\epsilon||g|_{B(\Omega)} - 2\epsilon(K^*w, g+h) + \epsilon^2\|K(g+h)\|^2_{L_2(\Omega)} + \gamma|\epsilon|\|h\|_{H^{-1}(\Omega)} \geq 0.$$

Dividing both sides of the latter equation by $\epsilon > 0$, we obtain

$$-2(K^*w, g+h) + \epsilon\|K(g+h)\|^2_{L_2(\Omega)} + 2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)} \geq 0.$$

Taking $\epsilon \to 0$, yields

$$2(K^*w, g+h) \leq 2\alpha|g|_{B(\Omega)} + \gamma\|h\|_{H^{-1}(\Omega)}, \text{ for all } g \in B(\Omega), \ h \in H^{-1}(\Omega) \cap L_2(\Omega)$$

and thus,

$$\|K^*w\|_* \leq \frac{1}{2}. \tag{5.8}$$

If we take $-1 < \epsilon < 0$, and replace (h, g) with (v, u) in equation (5.7), then (5.7) implies

$$2\epsilon(K^*w, u+v) \leq \epsilon\left(2\alpha|u|_{B(\Omega)} + \gamma\|v\|_{H^{-1}(\Omega)}\right) + \epsilon^2\|K(u+v)\|^2_{L_2(\Omega)}. \tag{5.9}$$

Dividing again by $\epsilon < 0$, we obtain $2(K^*w, u+v) \geq \left(2\alpha|u|_{B(\Omega)} + \gamma\|v\|_{H^{-1}(\Omega)}\right)$. Therefore equality holds,

$$(K^*w, u+v) = \frac{1}{2}\left(2\alpha|u|_{B(\Omega)} + \gamma\|v\|_{H^{-1}(\Omega)}\right), \tag{5.10}$$

and (5.10) together with (5.8) implies $\|K^*w\|_* = \frac{1}{2}$.

Conversely, if (5.4) holds for some (v, u) and $\|K^*w\|_* = \frac{1}{2}$, then for any $g \in B(\Omega)$, $h \in H^{-1}(\Omega) \cap L_2(\Omega)$,

$$2\alpha|u + \epsilon g|_{B(\Omega)} + \|w - \epsilon K(g+h)\|^2_{L_2} + \gamma\|v + \epsilon h\|_{H^{-1}(\Omega)}$$
$$\geq \ 2(K^*w, u + \epsilon g + v + \epsilon h) + \|w\|^2_{L_2(\Omega)} - 2\epsilon(K^*w, g+h) + \epsilon^2\|K(g+h)\|^2_{L_2(\Omega)}$$
$$= \ 2(K^*w, u+v) + \|w\|^2_{L_2(\Omega)} + \epsilon^2\|K(g+h)\|^2_{L_2(\Omega)}$$
$$= \ 2\alpha|u|_{B(\Omega)} + \gamma\|v\|_{H^{-1}(\Omega)} + \|w\|^2_{L_2(\Omega)} + \epsilon^2\|K(g+h)\|^2_{L_2(\Omega)}$$
$$\geq \ 2\alpha|u|_{B(\Omega)} + \gamma\|v\|_{H^{-1}(\Omega)} + \|w\|^2_{L_2(\Omega)} = \mathcal{G}_f(v, u) \ .$$

Therefore, (v, u) is an optimal decomposition and minimizer of (5.3). $\qquad\square$

5.2 Numerical Illustration

We present in Figure 14 a numerical result obtained with the proposed hybrid approach for the case $K = I$. The proposed scheme is numerically stable and faster than the method proposed in (Osher et al., 2003), where the minimization was solved by a 4th-order non-linear PDE with restrictive CFL condition. The method is also simpler than the method from (Vese and Osher, 2003) with $p = 2$. As expected, we can see that the proposed method gives better cartoon and texture separation than by the ROF method (Rudin et al., 1992) (corresponding to $\gamma = 0$, $v = 0$).

Acknowledgement

References

Ambrosio, L., N. Fusco, and D. Pallara, 2000: *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, Oxford.

Anthoine, S., 2005: Different wavelet-based approaches for the separation of noisy and blurred mixtures of components. application to astrophysical data. *PhD Thesis*.

Aubert, G. and J.-F. Aujol, 2005: Modeling very oscillating signals. Applications to image processing. *Applied Mathematics and Optimization*, **51 (2)**, 163.

Aubert, G. and L. Vese, 1997: A variational method in image recovery. *SINUM*, **34 (5)**, 1948–1979.

Bect, J., L. Blanc-Féraud, G. Aubert, and A. Chambolle, 2004: A l(1)-unified variational framework for image restoration. *Lecture Notes in Computer Science*, **2034**, 1–13.

Boyd, S. and L. Vandenberghe, 2005: *Convex Optimization*. Cambridge University Press, Cambridge.

Chambolle, A., 2004: An Algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, **20 (1-2)**, 89.

Chambolle, A. and P.-L. Lions, 1997: Image recovery via total variation minimization and related problems. *Numerische Mathematik*, **76 (2)**, 167.
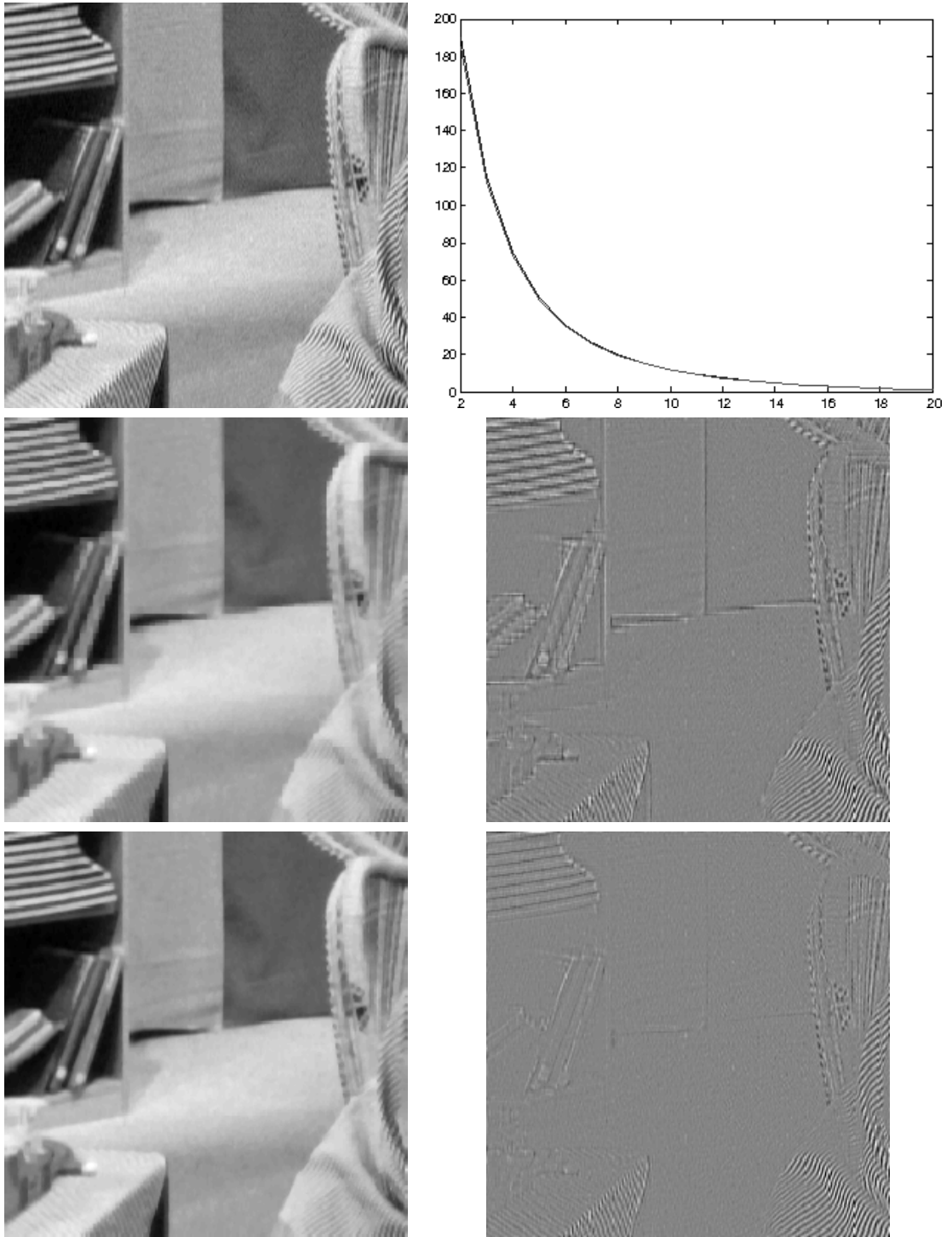
Figure 14: Image decomposition into cartoon and texture by the combined wavelet-PDE scheme, with parameters $J = 6$, $\gamma = 0.01$, $\alpha = 3$, 20 iterations. Top: original image f (left) and plot of $\|u^{n+1} - u^n\|$, $\|v^{n+1} - v^n\|$. Middle: u and $f - u$ obtained by the PDE scheme only ($\gamma = 0$, $v = 0$). Bottom: u and v obtained by the combined scheme.

Chen, S., D. Donoho, and M. Saunders, 1999: Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, **20 (1)**, 33–61.

Cohen, A., I. Daubechies, and J.-C. Feauveau, 1992: Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, **45**, 485–560.

Cohen, A., R. DeVore, P. Petrushev, and H. Xu, 1999: Nonlinear Approximation and the Space $BV(\mathbb{R}^2)$. *American Journal of Mathematics*, **(121)**, 587–628.

Coifman, R. and D. Donoho, 1995: Translation–invariant de–noising. in *Wavelets and Statistics, A. Antoniadis and G. Oppenheim,* eds., Springer–Verlag, New York, 125–150.

Combettes, P. L. and V. R. Wajs, 2005: Signal recovery by proximal forward–backward splitting. *Multiscale Modeling and Simulation*, **4 (4)**, 1168.

Dahmen, W., 1996: Stability of multiscale transformations. *The Journal of Fourier Analysis and Applications*, **2**, 341–361.

Daubechies, I., 1992: *Ten Lectures on Wavelets.* SIAM, Philadelphia.

———, 1993: Wavelet transforms and orthonormal wavelet bases. *Proceedings of Symposia in Applied Mathematics*, **(47)**.

Daubechies, I., M. Defrise, and C. DeMol, 2004: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Maths.*, **57**, 1413–1541.

Daubechies, I. and G. Teschke, 2004: Wavelet–based image decomposition by variational functionals. *Proc. SPIE Vol. 5266, p. 94-105, Wavelet Applications in Industrial Processing; Frederic Truchetet; Ed.*

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

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

Daudet, L. and B. Torrsani, 2002: Hybrid representations for audiophonic signal encoding. *Signal Processing, Special issue on Image and Video Coding Beyond Standards*, **82 (11)**, 1595–1617.

Defrise, M. and C. DeMol, 2004a: Inverse imaging with mixed penalties. *Conference proceedings.*

———, 2004b: Linear inverse problems with mixed smoothness and sparsity constraints. *Preprint.*

Demengel, F. and R. Temam, 1984: Convex functions of a measure and applications. *Indiana University Mathematics Journal*, **33 (5)**, 673.

DeVore, R., B. Jawerth, and V. Popov, 1988: Interpolation of besov spaces. *Trans. Math. Soc.*, **305**, 397–414.

——, 1992: Compression of wavelet decompositions. *Amer. J. Math.*, **114**, 737–785.

Ekeland, I. and R. Témam, 1999: *Convex Analysis and Variational Problems (Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics, Classics in Applied Mathematics, Philadelphia.

Elad, M., J.-L. Starck, P. Querre, and D. Donoho, 2005: Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis.*

Evans, L. and R. Gariepy, 1991: *Measure Theory and Fine Properties of Functions.* CRC Press.

Fernandes, F. C. A., R. v. Spaendonck, M. J. Coates, and S. Burrus, 2000: Directional Complex-Wavelet Processing. *Proceedings of Society of Photo-Optical Instrumental Engineers—SPIE2000, Wavelet Applications in Signal Processing VIII,* San Diego.

Fornasier, M. and H. Rauhut, 2006: Recovery algorithms for vector valued data with joint sparsity constraint. *Preprint.*

Frazier, M. and B. Jawerth, 1990: A discrete transform and decompositions of distribution spaces. *J. of Functional Anal.*, **93**, 34–170.

Geman, D. and G. Reynolds, 1992: Constrained restoration and the recovery of discontinuities. *IEEE Transactions on PAMI*, **14 (3)**, 367–383.

Geman, S. and D. Geman, 1984: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on PAMI*, **6 (6)**, 721–741.

Hiriart-Urruty, J.-B. and C. Lemaréchal, 1993: *Convex Analysis and Minimization Algorithms I: Fundamentals.* Springer, Berlin.

Jaillet, F. and B. Torresani, 2005: Time–frequency jigsaw puzzle: adaptive multiwindow and multilayered Gabor expansions. *preprint.*

Kinsbury, N., 1999: Image processing with complex wavelets. *Phil. Trans. R. Soc. Lond.*

Le, T. and L. Vese, 2005: Image Decomposition Using Total Variation and $div(BMO)$. *SIAM Multiscale Modeling and Simulation*, **4 (2)**, 390–423.

Mallat, S. and S. Zhong, 1992: Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14 (7)**, 710–732.

Meyer, Y., 2002: Oscillating Patterns in Image Processing and Nonlinear Evolution Equations. *University Lecture Series Volume 22, AMS.*

Molla, S. and B. Torresani, 2005: A hybrid scheme for encoding audio signal using hidden Markov models of waveforms. *Applied and Computational Harmonic Analysis*, **18 (2)**, 137–166.

Opial, Z., 1967: Weak convergence of the sequence of successive approximations for non-expansive mappings. *Bull. Amer. Math. Soc.*, **73 (4)**, 591–597.

Osher, S., A. Solé, and L. Vese, 2003: Image decomposition and restoration using total variation minimization and the H^{-1} norm. *SIAM Multiscale Modeling and Simulation*, **1 (3)**, 349.

Perona, P. and J. Malik, 1990: Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, **12**, 629–639.

Rockafellar, R. T. and R. J.-B. Wets, 1998: *Variational Analysis*. Springer, Berlin.

Rudin, L. and S. Osher, 1994: Total variation based image restoration with free local constraints . *Proceedings*, **ICIP-1994**, 31–35.

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

Selesnick, I. W., 2001: Hilbert transform pairs of wavelet bases. *IEEE Signal Processing Letters*, **8 (6)**, 170–173.

Starck, J.-L., M. Elad, and D. Donoho, 2005: Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transactions on Image Processing*, **14 (10)**, 1570–1582.

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

Triebel, H., 1978: *Interpolation Theory, Function Spaces, Differential Operators*. Verlag der Wissenschaften, Berlin.

Vese, L., 2001: A study in the BV space of a denoising-deblurring variational problem. *Applied Mathematics and Optimization*, **44 (2)**, 131–161.

Vese, L. and S. Osher, 2003: Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, **19 (1-3)**, 553–572.

———, 2004: Image denoising and decomposition with total variation minimization and oscillatory functions. *Journal of Mathematical Imaging and Vision*, **20 (1-2)**, 7–18.