International Journal of Wavelets, Multiresolution and Information Processing
© World Scientific Publishing Company

WAVELET-CONSTRAINED IMAGE RESTORATION

PATRICK L. COMBETTES

Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6
75005 Paris, France
plc@math.jussieu.fr

JEAN-CHRISTOPHE PESQUET

Institut Gaspard Monge and CNRS UMR 8049, Université de Marne la Vallée
77454 Marne la Vallée Cedex 2, France
pesquet@univ-mlv.fr

Received (Day Month Year)
Revised (Day Month Year)
Communicated by (xxxxxxxxxx)

Image restoration problems can naturally be cast as constrained convex programming problems in which the constraints arise from *a priori* information and the observation of signals physically related to the image to be recovered. In this paper, the focus is placed on the construction of constraints based on wavelet representations. Using a mix of statistical and convex-analytical tools, we propose a general framework to construct wavelet-based constraints. The resulting optimization problem is then solved with a block-iterative parallel algorithm which offers great flexibility in terms of implementation. Numerical results illustrate an application of the proposed framework.

Keywords: convex programming; image restoration; image denoising; Stein unbiased estimate; subgradient projection; wavelets.

AMS Subject Classification: XXX, XXX, XXX

1. Introduction

The classical linear restoration problem is to find the original form of an image $\overline{x}$ in a real Hilbert space $\mathcal{H}$ from the observation of a degraded image

$$y = L\overline{x} + u, \tag{1.1}$$

where $L\colon \mathcal{H} \to \mathcal{H}$ is a bounded linear operator modeling the blurring process and u models an additive noise perturbation ($L = \mathrm{Id}$ in denoising problems). Numerous approaches have been developed over the past three decades to solve this problem [1,8,16,29,32,35,36]. Despite their apparent disparity, these restoration problems can typically be posed as optimization problems in which an appropriate objective

function is minimized under certain constraints. Restricting ourselves to convex problems, the goal is therefore to

$$\text{find } \widehat{x} \in S = \bigcap_{i=1}^{m} S_i \quad \text{such that } J(\widehat{x}) = \inf J(S), \tag{1.2}$$

where the objective $J\colon \mathcal{H} \to]-\infty, +\infty]$ is a convex function and the constraint sets $(S_i)_{1\leq i\leq m}$ are convex subsets of $\mathcal{H}$. These constraints arise from *a priori* knowledge about the model (1.1) and the original image $\overline{x}$. Constraint sets can generally be represented as level sets, i.e.,

$$(\forall i \in \{1, \ldots, m\}) \;\; S_i = \mathrm{lev}_{\leq \delta_i} f_i, \tag{1.3}$$

where $f_i\colon \mathcal{H} \to]-\infty, +\infty]$ is a convex function and $\delta_i \in \mathbb{R}$. Examples of relevant functions f_i modeling spatial or spectral constraints can be found in [8,10,32].

In image denoising and restoration problems, the wavelet transform has been used in a variety of prescriptions [2,4,5,18,22]. The viewpoint adopted in the present work is that the wavelet transform can be exploited to construct various constraints on $\widehat{x}$ that can be used in conjunction with standard constraints to refine the feasibility set S in (1.2). Our general constraint model is described in Section 2, where we also review existing work. In Section 3, we introduce a new class of convex constraints arising from probabilistic information. These constraints are constructed via Stein's identity and turn out to be simple to handle numerically via projection methods. In Section 4, the parallel block-iterative convex programming framework of [9] is shown to provide a viable numerical scheme to solve the resulting optimization problem. We illustrate the proposed approach through a numerical image denoising application in Section 5.

2. Wavelet constraint model and previous work

2.1. *Notation*

In this paper, $\mathcal{H}$ will be either $L^2(\mathbb{R}^2)$ in analog models or $\ell^2(\mathbb{Z}^2)$ in discrete models. The scalar product and norm of $\mathcal{H}$ are denoted by $\langle \cdot \mid \cdot \rangle$ and $\|\cdot\|$, respectively. The adjoint of a bounded linear operator T is denoted by T^*. The level set of a function $f_i\colon \mathcal{H} \to \mathbb{R}$ at height $\delta_i \in \mathbb{R}$ is $\mathrm{lev}_{\leq \delta_i} f_i = \big\{x \in \mathcal{H} \mid f_i(x) \leq \delta_i\big\}$. The 2-D wavelet transform [26] in a separable wavelet basis $\mathcal{B}$ is denoted by $W^{\mathcal{B}}_{j,d}$, where $d \in \{1, 2, 3\}$ is the orientation parameter and $j \in \mathbb{Z}$ ($j \in \mathbb{Z}_-$ for discrete models) is the resolution level (here, coarser resolutions are obtained as $j \to -\infty$). The wavelet coefficients of $x \in \mathcal{H}$ are denoted by $\big(w^{\mathcal{B}}_{j,d,k}(x)\big)_{k\in\mathbb{Z}^2}$; in other words, $W^{\mathcal{B}}_{j,d}\colon \mathcal{H} \to \ell^2(\mathbb{Z}^2)\colon x \mapsto \big(w^{\mathcal{B}}_{j,d,k}(x)\big)_{k\in\mathbb{Z}^2}$. As usual, $\mathcal{N}(0,1)$ denotes a standard normal random variable. The characteristic function of a set Δ is denoted by 1_Δ. Finally, $\mathrm{sign}(t)$ takes value 1, 0, or -1, according as $t > 0$, $t = 0$, or $t < 0$.

2.2. *General constraint model*

A wavelet-based constraint function f_i in (1.3) can be constructed as follows. Let us fix a separable wavelet basis $\mathcal{B}$, a set $I \subset \mathbb{Z} \times \{1,2,3\}$ containing the indices of the retained scales j and directions d, and a convex function $\varphi\colon \ell^2(\mathbb{Z}^2) \to \,]-\infty,+\infty]$. We consider constraint functions of the form

$$f_i = \varphi \circ W_I^{\mathcal{B}} \tag{2.4}$$

or of the form

$$f_i = \varphi \circ W_I^{\mathcal{B}} \circ L, \tag{2.5}$$

where $W_I^{\mathcal{B}} = (W_{j,d}^{\mathcal{B}})_{(j,d)\in I}$. We thus obtain an inequality constraint on the wavelet coefficients of an image x itself or of its blurred version Lx.

2.3. *Previous work*

Wavelet constraints have already been used in the literature in specific contexts. An early development in that direction is [27], where the constraint imposed in the wavelet domain can be written as

$$(W^{\mathcal{B}}x)1_\Delta = r1_\Delta, \tag{2.6}$$

where $W^{\mathcal{B}}$ denotes the "entire" wavelet transform (i.e., $W^{\mathcal{B}} = W_I^{\mathcal{B}}$, with $I = \mathbb{Z} \times \{1,2,3\}$ for analog models and $I = \mathbb{Z}_- \times \{1,2,3\}$ for discrete models), r is a reference wavelet coefficient sequence, and the support region Δ corresponds to the location of local maxima. This constraint fits the convex inequality format described by (1.3) and (2.4), where $\varphi\colon z \mapsto \|(z-r)1_\Delta\|^2$ and $\delta_i = 0$. Related work can be found in [30,31]. The constraint model (2.6) also appears in [19,20], where Δ is defined through a thresholding operation.

Another type of wavelet constraint was considered in the multiple wavelet image denoising approach of [6] where upper bounds were imposed on the Besov norm of candidate solutions. The functions f_i are of the form

$$f_i\colon x \mapsto \|W^{\mathcal{B}}x\|_{b^s_{p,q}}^q \tag{2.7}$$

where, for sufficiently regular wavelets, the norm

$$\|\cdot\|_{b^s_{p,q}}\colon W^{\mathcal{B}}x \mapsto \left(\sum_j 2^{jsq}\Big(\sum_{k,d} |w^{\mathcal{B}}_{j,d,k}(x)|^p\Big)^{q/p}\right)^{1/q} \tag{2.8}$$

is known to be equivalent to the norm of the Besov space $B^\sigma_{p,q}$ when $p \geq 1$, $q \geq 1$, and $s = \sigma + 1 - 2/p$ [28]. A similar approach was adopted in [14] to solve one-dimensional regression problems. Related formulations are investigated in [4,15,25].

In the Bayesian maximum entropy framework described in [24] the constraints arise from upper bounds on the p-th ($p \geq 1$) order absolute moments of the wavelet

coefficient sequences. Such a constraint is therefore given by (2.4), where I is a singleton and

$$\varphi\colon (\gamma_k)_{k\in\mathbb{Z}^2} \mapsto \sum_k |\gamma_k|^p. \tag{2.9}$$

Note that, in practice the summations in (2.7) and (2.9) are performed over a finite range and, therefore, the associated function f_i is finite.

2.4. *Remarks*

To further improve the range of applications of wavelet-constrained approaches, the following directions should be explored.

- In certain simple scenarios, the solution to (1.2) can be computed in a straightforward fashion. For instance, this situation occurs when $J\colon x \mapsto \|x-y\|^2$ is minimized subject to the single Besov ball constraint (2.7), with $p=1$, or $q=1$, or $p=q=2$ [4,25]. However, in general settings, the resulting optimization problem (1.2) requires more sophisticated numerical techniques. In particular, fast and flexible algorithms with parallel processing capabilities should be adopted.
- The proposed framework should lend itself to the incorporation of a large number of constraints and, in particular, should allow for the combination of constraints arising in both the spatial and the wavelet domains.
- Critical to the effectiveness of a constraint set of type (1.3) is the determination of the bound δ_i. In some cases, such bounds are known *a priori* [8,37]. In other cases, they must be estimated from the data, which is a difficult task in general. Therefore, one should investigate classes of constraints amenable to reliable bound estimation methods.

This paper aims at addressing the above issues.

3. New constraints based on probabilistic information

3.1. *Introduction*

In order to refine the feasibility set in (1.2), one should incorporate as many convex constraints as the available *a priori* knowledge and the observed data allow. Usually, such constraints arise from information about $\overline{x}$ itself or about the noise process u [7,8,12,32,34,37]. In this section, we develop a procedure for constructing statistical constraints in the wavelet domain. In our approach, Stein's identity is used to estimate reliably a bound δ_i in (1.3).

3.2. *Constraint set construction via Stein's identity*

Our constraint construction scheme relies on the following fact, which is a consequence of Stein's identity [33]. Henceforth, ψ is a real-valued function defined on $\mathbb{R}$.

Proposition 3.1. *Suppose that A and B are real-valued random variables such that*

(i) $\mathsf{E}|A|^2 < +\infty$;
(ii) $B - A$ *is Gaussian with mean zero and variance* σ^2;
(iii) A *and* $B - A$ *are independent;*
(iv) ψ *is continuous, piecewise differentiable, and*

$$(\forall \theta \in \mathbb{R}) \lim_{|t|\to+\infty} \psi(t) \exp\Big(-\frac{(t-\theta)^2}{2\sigma^2}\Big) = 0; \tag{3.10}$$

(v) $\mathsf{E}|\psi(B)|^2 < +\infty$ *and* $\mathsf{E}|\psi'(B)| < +\infty$.

Then $\mathsf{E}(A\psi(B)) = \mathsf{E}(B\psi(B)) - \sigma^2 \mathsf{E}\psi'(B)$.

Throughout, $\mathcal{B}$ is a wavelet basis of $\mathcal{H}$. The decomposition of the data formation model (1.1) in this basis is

$$(\forall j \in \mathbb{Z})(\forall d \in \{1,2,3\})(\forall k \in \mathbb{Z}^2) \;\; w^{\mathcal{B}}_{j,d,k}(y) = w^{\mathcal{B}}_{j,d,k}(L\overline{x}) + w^{\mathcal{B}}_{j,d,k}(u). \tag{3.11}$$

Our standing assumptions are as follows.

Assumption 3.1. For every $j \in \mathbb{Z}$ and $d \in \{1,2,3\}$, the following conditions are satisfied.

(i) u and $L\overline{x}$ are independent random processes.
(ii) The random variables $\big(w^{\mathcal{B}}_{j,d,k}(u)\big)_{k\in\mathbb{Z}^2}$ are independent and all distributed as a zero-mean Gaussian random variable $w^{\mathcal{B}}_{j,d}(u)$ with standard deviation $\sigma^{\mathcal{B}}_{j,d}$.
(iii) The random variables $\big(w^{\mathcal{B}}_{j,d,k}(L\overline{x})\big)_{k\in\mathbb{Z}^2}$ are independent and all distributed as a random variable $w^{\mathcal{B}}_{j,d}(L\overline{x})$ with finite variance.
(iv) The function ψ is continuous, piecewise differentiable, and

$$(\forall \theta \in \mathbb{R}) \quad \lim_{|t|\to+\infty} |t|\psi(t)^2 \exp\Big(-\frac{(t-\theta)^2}{2(\sigma^{\mathcal{B}}_{j,d})^2}\Big) = 0. \tag{3.12}$$

(v) $0 < \mathsf{E}\big|\psi\big(w^{\mathcal{B}}_{j,d}(y)\big)\big|^2 < +\infty$ and $\mathsf{E}\big|\psi'\big(w^{\mathcal{B}}_{j,d}(y)\big)\big|^2 < +\infty$, where

$$w^{\mathcal{B}}_{j,d}(y) = w^{\mathcal{B}}_{j,d}(L\overline{x}) + w^{\mathcal{B}}_{j,d}(u). \tag{3.13}$$

Remark 3.1. In image processing applications, the underlying spatial homogeneity assumption on the noise in (ii) is quite standard and physically founded. In the case of discrete models, if u is a zero mean i.i.d. Gaussian noise with standard deviation σ, then Assumption 3.1(ii) is satisfied with $\sigma^{\mathcal{B}}_{j,d} = \sigma$.

Let us now fix a resolution level $j \in \mathbb{Z}$ ($j \in \mathbb{Z}_-$ for discrete models) and an orientation $d \in \{1,2,3\}$. Moreover, for every $x \in \mathcal{H}$, let us define

$$T^{\mathcal{B}}_{j,d}(x) = \mathsf{E}\Big(w^{\mathcal{B}}_{j,d}(Lx)\psi(w^{\mathcal{B}}_{j,d}(y)) - w^{\mathcal{B}}_{j,d}(y)\psi(w^{\mathcal{B}}_{j,d}(y)) + (\sigma^{\mathcal{B}}_{j,d})^2\psi'(w^{\mathcal{B}}_{j,d}(y))\Big). \tag{3.14}$$

The construction of the constraint will hinge on the following property.

Proposition 3.2. $T_{j,d}^{\mathcal{B}}(\overline{x}) = 0$.

Proof. Let us first observe that (3.12) implies (3.10). Indeed, let us fix $\theta \in \mathbb{R}$. Then, for $|t|$ large enough,

$$\left|\psi(t)\exp\Big(-\frac{(t-\theta)^2}{2\sigma^2}\Big)\right|^2 \leq |t|\psi(t)^2\exp\Big(-\frac{(t-\theta)^2}{2\sigma^2}\Big) \to 0. \tag{3.15}$$

Now set $A = w_{j,d}^{\mathcal{B}}(L\overline{x})$ and $B = w_{j,d}^{\mathcal{B}}(y)$. Then it follows from Assumption 3.1 that all the properties required in Proposition 3.1 are satisfied. Therefore,

$$\mathsf{E}\big(w_{j,d}^{\mathcal{B}}(L\overline{x})\psi(w_{j,d}^{\mathcal{B}}(y))\big) = \mathsf{E}\big(w_{j,d}^{\mathcal{B}}(y)\psi(w_{j,d}^{\mathcal{B}}(y))\big) - (\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}\psi'(w_{j,d}^{\mathcal{B}}(y)), \tag{3.16}$$

and the claim is proved. □

The conceptual constraint $T_{j,d}^{\mathcal{B}}(\overline{x}) = 0$ is not enforceable since the expectation is not tractable. As a result, it must be replaced by the practical constraint

$$T_{j,d}^{\mathcal{B},K_j}(\overline{x}) \approx 0, \tag{3.17}$$

where $T_{j,d}^{\mathcal{B},K_j}(\overline{x})$ is a consistent estimate of $T_{j,d}^{\mathcal{B}}(\overline{x})$ computed from a K_j^2-point observation window $\mathbb{K}_j = \{0,\dots,K_j-1\}^2$ in the wavelet domain. For every $x \in \mathcal{H}$, let us define

$$\begin{aligned}(\forall k \in \mathbb{Z}^2)\ \ z_k(x) = {} & w_{j,d,k}^{\mathcal{B}}(Lx)\psi(w_{j,d,k}^{\mathcal{B}}(y)) - w_{j,d,k}^{\mathcal{B}}(y)\psi(w_{j,d,k}^{\mathcal{B}}(y)) \\ & + (\sigma_{j,d}^{\mathcal{B}})^2\psi'(w_{j,d,k}^{\mathcal{B}}(y))\end{aligned} \tag{3.18}$$

and

$$T_{j,d}^{\mathcal{B},K_j}(x) = \frac{1}{K_j^2}\sum_{k\in\mathbb{K}_j} z_k(x). \tag{3.19}$$

We now show that the empirical estimate $T_{j,d}^{\mathcal{B},K_j}(\overline{x})$ is strongly consistent.

Proposition 3.3. *The random variables $(z_k(\overline{x}))_{k\in\mathbb{Z}^2}$ are independent and all distributed as the random variable*

$$z(\overline{x}) = w_{j,d}^{\mathcal{B}}(L\overline{x})\psi(w_{j,d}^{\mathcal{B}}(y)) - w_{j,d}^{\mathcal{B}}(y)\psi(w_{j,d}^{\mathcal{B}}(y)) + (\sigma_{j,d}^{\mathcal{B}})^2\psi'(w_{j,d}^{\mathcal{B}}(y)). \tag{3.20}$$

Moreover, $\mathsf{E}z(\overline{x}) = 0$ and

$$\mathsf{E}|z(\overline{x})|^2 = (\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}|\psi(w_{j,d}^{\mathcal{B}}(y))|^2 + (\sigma_{j,d}^{\mathcal{B}})^4\mathsf{E}|\psi'(w_{j,d}^{\mathcal{B}}(y))|^2 < +\infty. \tag{3.21}$$

Proof. The first claim follows from Assumption 3.1, whereas the identity $\mathsf{E}z(\overline{x}) = 0$ follows from Proposition 3.2. Furthermore,

$$\begin{aligned}\mathsf{E}|z(\overline{x})|^2 &= \mathsf{E}\big|\big(w_{j,d}^{\mathcal{B}}(y) - w_{j,d}^{\mathcal{B}}(L\overline{x})\big)\psi(w_{j,d}^{\mathcal{B}}(y)) - (\sigma_{j,d}^{\mathcal{B}})^2\psi'(w_{j,d}^{\mathcal{B}}(y))\big|^2 \\ &= \mathsf{E}|w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(y)) - (\sigma_{j,d}^{\mathcal{B}})^2\psi'(w_{j,d}^{\mathcal{B}}(y))|^2 \\ &= \mathsf{E}|w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(y))|^2 - 2(\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}\big(w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(y))\psi'(w_{j,d}^{\mathcal{B}}(y))\big) \\ &\quad + (\sigma_{j,d}^{\mathcal{B}})^4\mathsf{E}|\psi'(w_{j,d}^{\mathcal{B}}(y))|^2.\end{aligned} \tag{3.22}$$

Per Assumption 3.1(ii), $w_{j,d}^{\mathcal{B}}(u)$ is Gaussian with mean zero and standard deviation $\sigma_{j,d}^{\mathcal{B}}$. Let g be its density and let $h = w_{j,d}^{\mathcal{B}}(L\overline{x})$. Hence, $w_{j,d}^{\mathcal{B}}(y) = w_{j,d}^{\mathcal{B}}(u) + h$ and we obtain

$$\begin{aligned}\mathsf{E}|w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(y))|^2 &= \mathsf{E}\Big(\mathsf{E}\big(|w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(u)+h)|^2 \mid h\big)\Big)\\ &= \mathsf{E}\int_{\mathbb{R}} t^2\psi(t+h)^2 g(t)dt \end{aligned} \tag{3.23}$$

However, in view of Assumption 3.1(iv), an integration by parts yields

$$\begin{aligned}\int_{\mathbb{R}} t^2\psi(t+h)^2 g(t)dt &= -(\sigma_{j,d}^{\mathcal{B}})^2\int_{\mathbb{R}} t\psi(t+h)^2 dg(t)\\ &= (\sigma_{j,d}^{\mathcal{B}})^2\int_{\mathbb{R}} g(t)d\big(t\psi(t+h)^2\big)\\ &= (\sigma_{j,d}^{\mathcal{B}})^2\int_{\mathbb{R}} g(t)\big(\psi(t+h)^2 + 2t\psi(t+h)\psi'(t+h)\big)dt\\ &= (\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}\big(\psi(w_{j,d}^{\mathcal{B}}(u)+h)^2 \mid h\big)\\ &\quad + 2(\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}\big(w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(u)+h)\psi'(w_{j,d}^{\mathcal{B}}(u)+h) \mid h\big).\end{aligned}$$

Thus, we derive from (3.23) that

$$\begin{aligned}\mathsf{E}|w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(y))|^2 &= (\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}|\psi(w_{j,d}^{\mathcal{B}}(y))|^2\\ &\quad + 2(\sigma_{j,d}^{\mathcal{B}})^2\mathsf{E}\big(w_{j,d}^{\mathcal{B}}(u)\psi(w_{j,d}^{\mathcal{B}}(y))\psi'(w_{j,d}^{\mathcal{B}}(y))\big).\end{aligned} \tag{3.24}$$

In turn, combining this identity with (3.22) yields (3.21). □

A straightforward application of the strong law of large numbers now furnishes the announced strong consistency result.

Proposition 3.4. *$T_{j,d}^{\mathcal{B},K_j}(\overline{x}) \xrightarrow{\text{a.s.}} 0$ as $K_j \to +\infty$.*

In order to give (3.17) a precise statistical meaning, we need to investigate the asymptotic distribution of $T_{j,d}^{\mathcal{B},K_j}(\overline{x})$ as K_j becomes arbitrarily large. To this end, let us define

$$V_{j,d}^{\mathcal{B},K_j} = \frac{\sigma_{j,d}^{\mathcal{B}}}{K_j^2}\sqrt{\sum_{k\in\mathbb{K}_j}|\psi(w_{j,d,k}^{\mathcal{B}}(y))|^2 + (\sigma_{j,d}^{\mathcal{B}})^2\sum_{k\in\mathbb{K}_j}|\psi'(w_{j,d,k}^{\mathcal{B}}(y))|^2}. \tag{3.25}$$

In view of (3.21), $K_j V_{j,d}^{\mathcal{B},K_j}$ is therefore the empirical standard deviation of $z(\overline{x})$ based on the observation window $\mathbb{K}_j$.

Theorem 3.1. *$T_{j,d}^{\mathcal{B},K_j}(\overline{x})/V_{j,d}^{\mathcal{B},K_j} \xrightarrow{\text{d}} \mathcal{N}(0,1)$ as $K_j \to +\infty$.*

Proof. It follows from Proposition 3.3 and the standard central limit theorem that

$$K_j T_{j,d}^{\mathcal{B},K_j}(\overline{x})/\sqrt{\mathsf{E}|z(\overline{x})|^2} \xrightarrow{\text{d}} \mathcal{N}(0,1) \quad \text{as} \quad K_j \to +\infty. \tag{3.26}$$

On the other hand, it follows from Proposition 3.3 and the strong law of large numbers that $|K_j V_{j,d}^{\mathcal{B},K_j}|^2 \xrightarrow{\text{a.s.}} \mathsf{E}|z(\overline{x})|^2$ as $K_j \to +\infty$. Since $\mathsf{E}|z(\overline{x})|^2 \neq 0$ by Assumption 3.1(v) and (3.21), $V_{j,d}^{\mathcal{B},K_j} \neq 0$ for K_j large enough. Hence, we deduce from Proposition 6.3.4 in [3] that

$$\sqrt{\mathsf{E}|z(\overline{x})|^2}/(K_j V_{j,d}^{\mathcal{B},K_j}) \xrightarrow{\text{d}} 1. \tag{3.27}$$

We can now invoke Proposition 6.3.8(ii) in [3] to derive from (3.26) and (3.27) that $T_{j,d}^{\mathcal{B},K_j}(\overline{x})/V_{j,d}^{\mathcal{B},K_j} \xrightarrow{\text{d}} \mathcal{N}(0,1)$. □

Remark 3.2. The proof of the above theorem is based on the a.s. convergence of $K_j V_{j,d}^{\mathcal{B},K_j}$ toward the standard deviation of $z(\overline{x})$. This property still holds if, in (3.25), the variance $\sigma_{j,d}^{\mathcal{B}}$ is replaced by a consistent estimate $\sigma_{j,d}^{\mathcal{B},K_j}$, that is

$$\sigma_{j,d}^{\mathcal{B},K_j} \xrightarrow{\text{a.s.}} \sigma_{j,d}^{\mathcal{B}} \quad \text{as} \quad K_j \to +\infty. \tag{3.28}$$

This shows that our approach is applicable when the variance of the noise is unknown provided that it can be consistently estimated.

Remark 3.3. The conclusions of Proposition 3.4 and Theorem 3.1 are direct consequences of the strong law of large number and of the central limit theorem. Consequently, they remain valid under much weaker assumptions on $\big(w_{j,d,k}^{\mathcal{B}}(u)\big)_{k\in\mathbb{Z}^2}$ and $\big(w_{j,d,k}^{\mathcal{B}}(L\overline{x})\big)_{k\in\mathbb{Z}^2}$ than those adopted in Assumptions 3.1(ii) and (iii). In particular, by applying sharper asymptotic results (see for instance [23]), mixing assumptions on the wavelet coefficients of the image of interest make it possible to extend the scope of our results, while providing more realistic models for natural images.

The practical significance of Theorem 3.1 is the following. Let us fix a confidence level $\mathsf{p}_i \in \,]0,1[$ and let $\mathrm{erf}\colon \tau \mapsto (2/\sqrt{\pi})\int_0^\tau e^{-t^2}dt$ be the error function. Then, assuming that K_j is large enough so that the normal approximation is legitimate, $|T_{j,d}^{\mathcal{B},K_j}(\overline{x})/V_{j,d}^{\mathcal{B},K_j}|$ will not exceed $\sqrt{2}\,\mathrm{erf}^{-1}(\mathsf{p}_i)$ with probability p_i. Accordingly, for K_j sufficiently large, the true image $\overline{x}$ lies in the set

$$S_i = \Big\{ x \in \mathcal{H} \;\Big|\; \big|T_{j,d}^{\mathcal{B},K_j}(x)\big| \leq \sqrt{2} V_{j,d}^{\mathcal{B},K_j} \mathrm{erf}^{-1}(\mathsf{p}_i) \Big\}, \tag{3.29}$$

to within the confidence level p_i. Now put

$$\eta_i = \sum_{k\in\mathbb{K}_j} w_{j,d,k}^{\mathcal{B}}(y)\psi(w_{j,d,k}^{\mathcal{B}}(y)) - (\sigma_{j,d}^{\mathcal{B}})^2 \psi'(w_{j,d,k}^{\mathcal{B}}(y)). \tag{3.30}$$

Then the constraint set S_i is readily seen to fit the general format described by (1.3) and (2.5) with $I = \{(j,d)\}$,

$$\varphi\colon (\mu_k)_{k\in\mathbb{Z}^2} \mapsto \left| \sum_{k\in\mathbb{K}_j} \mu_k \psi(w_{j,d,k}^{\mathcal{B}}(y)) - \eta_i \right|, \tag{3.31}$$

and

$$\delta_i = \sqrt{2} K_j^2 V_{j,d}^{\mathcal{B},K_j} \mathrm{erf}^{-1}(\mathsf{p}_i). \tag{3.32}$$

In view of (3.25), the bound δ_i is computable entirely from the observed image y. Geometrically, S_i is simply described as a hyperslab, i.e., a set of the form $S_i = \{x \in \mathcal{H} \mid \beta_1 \leq \langle x \mid a \rangle \leq \beta_2\}$, where $a \in \mathcal{H}$ and $(\beta_1, \beta_2) \in \mathbb{R}^2$.

3.3. *Confidence analysis*

In Section 3.2, the constraint (3.17) has been enforced through a statistical test that led to the construction of the set S_i in (3.29). As a result, S_i is a confidence region which contains the original image $\overline{x}$ to within the confidence level p_i. In practice, several statistical constraints of this type can be used simultaneously, since one has the option of considering several wavelet bases $\mathcal{B}$, several resolution levels j, several directions d, and several functions ψ. Suppose that q sets $(S_i)_{1 \leq i \leq q}$ of type (3.29) are to be used. In order to guarantee the reliability of the solutions to problem (1.2), one should select the confidence levels $(\mathsf{p}_i)_{1 \leq i \leq q}$ so as to achieve a preset confidence level c on $\bigcap_{i=1}^{q} S_i$. Following the analysis of [11], we can invoke Bonferroni's inequality

$$\mathsf{c} \geq 1 - \sum_{i=1}^{q} (1 - \mathsf{p}_i) \tag{3.33}$$

to adjust the individual parameters $(\mathsf{p}_i)_{1 \leq i \leq q}$. For instance, if all these parameters are taken to be equal to p, then choosing

$$\mathsf{p} \geq 1 - (1 - \mathsf{c}')/q \tag{3.34}$$

will guarantee a global confidence level c on the intersection of the sets of at least c'.

3.4. *Choice of ψ*

Among the many possible choices for ψ, let us mention a few functions of interest.

- $\psi : t \mapsto |t|^{p-1}\mathrm{sign}(t)$, where $p > 1$. In this case, $\mathsf{E}\big(w^{\mathcal{B}}_{j,d}(y)\psi\big(w^{\mathcal{B}}_{j,d}(y)\big)\big) = \mathsf{E}|w^{\mathcal{B}}_{j,d}(y)|^p$, which is similar to the p-th order absolute moment constraint in (2.9) when $L = \mathrm{Id}$. In particular, when $p = 2$, $\psi : t \mapsto t$ and the associated set S_i corresponds to a constraint on the correlation between $w^{\mathcal{B}}_{j,d}(L\overline{x})$ and $w^{\mathcal{B}}_{j,d}(y)$. In this simple case, (3.16) reduces to
$$\mathsf{E}\big(w^{\mathcal{B}}_{j,d}(L\overline{x})w^{\mathcal{B}}_{j,d}(y)\big) = \mathsf{E}|w^{\mathcal{B}}_{j,d}(y)|^2 - (\sigma^{\mathcal{B}}_{j,d})^2 = \mathsf{E}|w^{\mathcal{B}}_{j,d}(L\overline{x})|^2 \tag{3.35}$$
and, therefore, a constraint is placed on the energy of the wavelet coefficients of Lx in a given subband (j, d).
- $\psi : t \mapsto \tanh(t/\alpha)$ where $\alpha \in \,]0, +\infty[$. This function provides a smooth approximation to the sign function. When $\alpha \downarrow 0$, $\mathsf{E}\big(w^{\mathcal{B}}_{j,d}(y)\psi(w^{\mathcal{B}}_{j,d}(y))\big) \rightarrow \mathsf{E}|w^{\mathcal{B}}_{j,d}(y)|$ and the associated constraint is akin to an absolute moment constraint.
- $\psi : t \mapsto t\big(\tanh((t+\chi)/\alpha) - \tanh((t-\chi)/\alpha)\big)/2$ where $\alpha \in \,]0, +\infty[$ and $\chi \in \,]0, +\infty[$. When $\alpha \downarrow 0$, $\psi(t) \rightarrow t1_{]-\chi,\chi[}(t)$. We thus enforce a constraint on the correlation between $w^{\mathcal{B}}_{j,d}(L\overline{x})$ and $w^{\mathcal{B}}_{j,d}(y)$ with absolute values below the

threshold χ. In practice, χ may be chosen as the so-called universal threshold [17]. In the same vein, another pertinent choice is $\psi\colon t \mapsto \tanh(t/\alpha)\big(\tanh((t+\chi)/\alpha) - \tanh((t-\chi)/\alpha)\big)/2$.

4. Numerical algorithm

4.1. *Subgradient projections*

We review a few basic facts, see [10] and the references therein for details.

Let S_i be a nonempty closed and convex subset of $\mathcal{H}$ and let x be a point in $\mathcal{H}$. Then there exists a unique point $P_i x \in S_i$, called the projection of x onto S_i, such that $\|x - P_i x\| = d_{S_i}(x)$, where $d_{S_i}(x) = \inf \|x - S_i\|$. Now suppose that S_i is given by (1.3), where f_i is continuous and convex (since $S_i = \operatorname{lev}_{\leq 0} d_{S_i}$, such a representation always exists). The subdifferential of f_i at $x \in \mathcal{H}$ is

$$\partial f_i(x) = \big\{g_i \in \mathcal{H} \mid (\forall z \in \mathcal{H})\ \ \langle z - x \mid g_i\rangle + f_i(x) \leq f_i(z)\big\}. \tag{4.36}$$

Now, fix $x \in \mathcal{H}$, a subgradient $g_i \in \partial f_i(x)$, and define

$$H_x = \begin{cases} \big\{z \in \mathcal{H} \mid \langle z - x \mid g_i\rangle \leq \delta_i - f_i(x)\big\}, & \text{if } \ f_i(x) > \delta_i; \\ \mathcal{H}, & \text{if } \ f_i(x) \leq \delta_i. \end{cases} \tag{4.37}$$

Then $S_i \subset H_x$ and the projection of x onto H_x, i.e.,

$$G_i x = P_{H_x} x = \begin{cases} x + \dfrac{\delta_i - f_i(x)}{\|g_i\|^2} g_i, & \text{if } \ f_i(x) > \delta_i; \\ x, & \text{if } \ f_i(x) \leq \delta_i, \end{cases} \tag{4.38}$$

is called a subgradient projection of x onto S_i. We note that computing $G_i x$ requires only the availability of a subgradient (a gradient in the differentiable case) of f_i at x and is therefore much more economical than computing the exact projection $P_i x$, as the latter amounts to solving a constrained quadratic minimization problem. However, when $P_i x$ is easy to compute, one can set $f_i = d_{S_i}$ and obtain $G_i x = P_i x$.

Now suppose that φ is (convex and) lower semicontinuous and that there exists a point $x \in \mathcal{H}$ such that φ is finite and continuous at $W_I^{\mathcal{B}} x$. Then, for the constraints described in (2.4) and (2.5), standard convex calculus yields [21]

$$\partial f_i(x) = (W_I^{\mathcal{B}})^* \partial\varphi(W_I^{\mathcal{B}} x) \tag{4.39}$$

and

$$\partial f_i(x) = L^* (W_I^{\mathcal{B}})^* \partial\varphi(W_I^{\mathcal{B}} L x), \tag{4.40}$$

respectively. When $\mathcal{B}$ is an orthonormal basis of $\mathcal{H}$, $(W_I^{\mathcal{B}})^*$ is directly related to the inverse wavelet transform since $(W^{\mathcal{B}})^* = (W^{\mathcal{B}})^{-1}$. For instance, if the Besov ball constraint defined through (2.7) is expressed in an orthonormal wavelet basis $\mathcal{B}$, a subgradient g_i of f_i at $x \in \mathcal{H}$ is obtained component-wise in the wavelet domain as

$$w_{j,d,k}^{\mathcal{B}}(g_i) = 2^{jsq} q \|W_{I_j}^{\mathcal{B}}\|_p^{q-p} |w_{j,d,k}^{\mathcal{B}}(x)|^{p-1} \operatorname{sign}(w_{j,d,k}^{\mathcal{B}}(x)), \tag{4.41}$$

where $I_j = \{j\} \times \{1,2,3\}$. This expression can then be fed back directly into (4.38) to obtain a subgradient projection. On the other hand, computing an exact projection in this case would require solving a constrained quadratic programming iteratively.

4.2. *Algorithm*

The objective is to solve our basic problem (1.2) efficiently under the constraint representation (1.3). To this end, we shall use the following block-iterative algorithm proposed in [9]. The convergence of this algorithm has been studied in a general setting in [9]. In order to avoid technical digressions, we state only the finite dimensional convergence result of interest in the context of the subsequent numerical simulations. Thus, images are assumed to be sampled on a finite $K \times K$ grid and the image space is therefore the Euclidean space $\mathbb{R}^N$, where $N = K^2$. Our detailed set of assumptions is as follows.

Assumption 4.1.

(i) $\mathcal{H} = \mathbb{R}^N$ and $(S_i)_{1\leq i\leq m}$ are defined by (1.3), where the functions $(f_i)_{1\leq i\leq m}$ are finite and convex, and $S = \bigcap_{i=1}^m S_i \neq \emptyset$.
(ii) $J\colon \mathbb{R}^N \to \,]-\infty,+\infty]$ is convex and lower semicontinuous.
(iii) There exists $z \in S$ such that $J(z) < +\infty$, $C = \mathrm{lev}_{\leq J(z)}\, J$ is bounded, and J is differentiable and strictly convex on C.

Let us observe that, under assumption (i), properties (ii) and (iii) hold in particular when $J\colon \mathbb{R}^N \to \mathbb{R}$ is strictly convex, differentiable, and coercive in the sense that $\lim_{\|x\|\to+\infty} J(x) = +\infty$.

Algorithm 4.1.

① Fix $\varepsilon \in \,]0, 1/m[$. Let x_0 be the minimizer of J over $\mathcal{H}$ and set $n = 0$.
② Take a nonempty index set $I_n \subset \{1,\ldots,m\}$.
③ For every $i \in I_n$, take $g_{i,n} \in \partial f_i(x_n)$ and compute the subgradient projection

$$p_{i,n} = \begin{cases} x_n + \dfrac{\delta_i - f_i(x_n)}{\|g_{i,n}\|^2} g_{i,n}, & \text{if } f_i(x_n) > \delta_i; \\ x_n, & \text{if } f_i(x_n) \leq \delta_i. \end{cases}$$

④ Set $z_n = x_n + \lambda_n \big(\sum_{i\in I_n} \omega_{i,n} p_{i,n} - x_n\big)$, where
(a) $(\omega_{i,n})_{i\in I_n}$ lies in $\,]\varepsilon, 1]$ and $\sum_{i\in I_n} \omega_{i,n} = 1$;
(b) $\lambda_n = \begin{cases} \dfrac{\sum_{i\in I_n} \omega_{i,n} \|p_{i,n} - x_n\|^2}{\big\|\sum_{i\in I_n} \omega_{i,n} p_{i,n} - x_n\big\|^2}, & \text{if } \max\limits_{i\in I_n} \big(f_i(x_n) - \delta_i\big) > 0; \\ 1, & \text{otherwise.} \end{cases}$

⑤ Set $\begin{cases} D_n = \big\{x \in \mathcal{H} \mid \langle x_n - x \mid \nabla J(x_n)\rangle \leq 0\big\} \\ H_n = \big\{x \in \mathcal{H} \mid \langle z_n - x \mid z_n - x_n\rangle \leq 0\big\}. \end{cases}$
⑥ Let x_{n+1} be the minimizer of J over $D_n \cap H_n$.

⑦ Set $n = n + 1$ and go to ②.

Theorem 4.1. [9] Suppose that Assumption 4.1 is satisfied and that there exists a strictly positive integer M such that, for every $i \in \{1, \ldots, m\}$ and every $n \in \mathbb{N}$, $i \in \bigcup_{k=n}^{n+M-1} I_k$. Then every sequence generated by Algorithm 4.1 converges to the unique solution $\widehat{x}$ to (1.2).

4.3. *Remarks*

To determine the update x_{n+1} at iteration n of Algorithm 4.1, one first computes *simultaneously* the subgradient projections $(p_{i,n})_{i\in I_n}$ of x_n onto the selected sets $(S_i)_{i\in I_n}$ and then forms a relaxed convex average of these projections. This step can be decomposed on a parallel architecture by assigning the computation of a subgradient projection to each processor. The next important step is ⑥, which amounts to minimizing J over the intersection of (at most) two half-spaces. This is a simple convex program with (one or) two affine constraints, and it can sometimes be solved in closed form [9,10].

The algorithm offers great flexibility in the choice of the constraints to be activated at each iteration. It is therefore possible to match the computational load of each iteration with the parallel architecture at hand. In particular, constraints expressed in several domains (e.g., time, frequency, and several wavelet domains) can be processed independently and concurrently. Added flexibility is supplied by the fact that the weights and the relaxations can vary at each iteration.

As discussed in [9,10], although no convergence rate can be computed for this type of nonlinear programming method in general, it displays nice convergence patterns due to the deep cuts induced by the surrogate half-space H_n at Step ⑤ (see Fig. 8 for an illustration). Moreover, in terms of stopping rule, feasibility with respect to the constraints can be used since x_n is the solution $\widehat{x}$ to (1.2) if and only if $x_n \in S$; see Proposition 3.1(v) in [9].

5. Numerical example

The image y of Fig. 2 is obtained by adding i.i.d. zero mean Gaussian noise u to the 256×256 8-bit Lena image $\overline{x}$ shown in Fig. 1. The image space is therefore $\mathcal{H} = \mathbb{R}^N$, where $N = 256^2$. The mean-square error (MSE) between the original and the degraded image is 900 (the signal-to-noise ratio is 11.92 dB). The variance of the noise is assumed to be known. The constraint set arising from the knowledge of the pixel range values is $S_1 = [0, 255]^N$. The objective function is

$$J\colon x \mapsto \|x - y\|^2, \tag{5.42}$$

which corresponds to the classical maximum likelihood criterion.

Different wavelet-based constraints are evaluated considering two separable orthonormal wavelet representations (using symlets [13] of length 8 and 12) which are performed over 4 resolution levels j. The resulting optimization problem is solved

with Algorithm 4.1 (see [10] for the closed-form implementation of Step ⑥ allowed by (5.42)).

In the first experiment the sets S_2 and S_3 are the Besov balls arising from (2.7), where $p = 1.1$, $q = 1.2$, and $s = 0.7$. The wavelet basis $\mathcal{B}$ used in the construction of these sets is the length-8 symlet basis for the set S_2, and the length-12 symlet basis for the set S_3. We observe that no efficient solution technique has been proposed for this particular setting in the literature. The image displayed in Fig. 3 is obtained with the exact Besov bounds computed directly from the original image. Next, we show in Figs. 4 and 5 the images obtained when the radii of the Besov balls are underestimated by a factor 0.2 and overestimated by a factor 3, respectively. It is clear that the quality of the restoration depends strongly on the reliable estimation of the radii which, in the presence of noisy data, constitutes a challenging problem.

The second experiment focuses on the proposed approach with the constraint sets of (3.29). Besov ball constraints are not involved in this experiment. Based on the discussion of Section 3.4, three functions ψ are considered, namely

- $\psi\colon t \mapsto \tanh(t/a)$,
- $\psi\colon t \mapsto t\big(\tanh((t+\chi)/a) - \tanh((t-\chi)/a)\big)$,
- $\psi\colon t \mapsto \tanh(t/a)\big(\tanh((t+\chi)/a) - \tanh((t-\chi)/a)\big)$,

where $a = 10$ and $\chi = 100$. Since 12 subbands $(j,d) \in \{-1,-2,-3,-4\} \times \{1,2,3\}$ are used for the detail coefficients, we have a total of $3 \times 4 \times 3 = 36$ sets $(S_i)_{2 \le i \le 37}$. We choose a fixed confidence level p on these sets. The value of p is derived from (3.34) so as to achieve a global confidence level of at least $\mathsf{c}' = 0.8$. As pointed out in Section 3.2, the parameters defining the sets $(S_i)_{2 \le i \le 37}$ are entirely determined from the data y. The denoised image is shown in Fig. 6. For comparison purposes, the image produced by the SUREshrink thresholding method [18] with a length-8 symlet basis is shown in Fig. 7. An inspection of the last two figures and of the values of the mean square errors reveals that the proposed wavelet-based constraints lead to an improvement in the quality of the recovery. It is worth mentioning that using only the length-8 symlet basis in this experiment would lead to a lower performance (MSE = 175.6) that would however still outperform the SUREshrink approach. Let us also point out that, since the sample sizes used to compute the empirical statistics is relatively large, the MSE displays little sensitivity with respect to the choice of the confidence level c' in a neighborhood of the level 0.8 used in this experiment. Thus, the MSE varies monotonically from 149.46 for $\mathsf{c}' = 0.7$ to 150.05 for $\mathsf{c}' = 0.9$. Finally, from a numerical standpoint, the method has been found to converge rapidly. For instance, the decibel value of the normalized mean square error pattern in the case of the restoration obtained in Fig. 6 is plotted in Fig. 8.

References

1. H. C. Andrews and B. R. Hunt, *Digital Image Restoration* (Prentice-Hall, 1977).
2. A. Antoniadis, D. Leporini, and J.-C. Pesquet, *Wavelet thresholding for some classes of non-Gaussian noise, Statist. Neerlandica,* **56** (2002) 434–453.

3. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods,* 2nd ed. (Springer, 1991).
4. A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage, IEEE Trans. Image Process.,* **7** (1998) 319–335.
5. S. G. Chang, B. Yu, and M. Vetterli, *Spatially adaptive wavelet thresholding with context modelling for image denoising*, *IEEE Trans. Image Process.,* **9** (2000) 1522–1531.
6. H. Choi and R. Baraniuk, *Multiple basis wavelet denoising using Besov projections, Proc. IEEE Int. Conf. Image Process.,* (1999) 595–599.
7. P. L. Combettes, *The foundations of set theoretic estimation, Proc. IEEE,* **81** (1993) 182–208.
8. P. L. Combettes, *The convex feasibility problem in image recovery*, in *Advances in Imaging and Electron Physics*, **95** (Academic Press, 1996) 155–270.
9. P. L. Combettes, *Strong convergence of block-iterative outer approximation methods for convex optimization, SIAM J. Control Optim.,* **38** (2000) 538–565.
10. P. L. Combettes, *A block-iterative surrogate constraint splitting method for quadratic signal recovery, IEEE Trans. Signal Process.,* **51** (2003) 1771–1782.
11. P. L. Combettes and T. J. Chaussalet, *Combining statistical information in set theoretic estimation, IEEE Signal Process. Lett.*, **3** (1996) 61–62.
12. P. L. Combettes and H. J. Trussell, *The use of noise properties in set theoretic estimation, IEEE Trans. Signal Process.,* **39** (1991) 1630–1641.
13. I. Daubechies, *Ten Lectures on Wavelets.* (SIAM, 1992).
14. L. T. Dechevsky, J. O. Ramsay, and S. I. Penev, *Penalized wavelet estimation with Besov regularity constraints, Math. Balkanica (N.S.)*, **13** (1999), 257–376.
15. C. de Mol and M. Defrise, *A note on wavelet-based inversion algorithms, Contemp. Math.*, **313** (2002) 85–96.
16. G. Demoment, *Image reconstruction and restoration: Overview of common estimation structures and problems, IEEE Trans. Acoust., Speech, Signal Process.,* **37** (1989) 2024–2036.
17. D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage, Biometrika*, **81** (1994) 425–455.
18. D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Stat. Assoc.*, **90** (1995) 1200–1224.
19. S. Durand and J. Froment, *Reconstruction of wavelet coefficients using total variation minimization, SIAM J. Sci. Comput.* **24** (2003) 1754–1767.
20. S. Durand and M. Nikolova, *Restoration of wavelet coefficients by minimizing a specially designed objective function, Proc. IEEE Int. Conf. Comput. Vision* (2003) 145–152.
21. I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationnels* (Dunod, 1974); *Convex Analysis and Variational Problems* (SIAM, 1999).
22. J. Kalifa and S. Mallat, *Minimax restoration and deconvolution*, in *Bayesian Inference in Wavelet-Based Models*, (Müller & Vidakovic, Eds.), Lect. Notes Stat. (Springer, 1999) 115–138.
23. X. Guyon, *Champs Aléatoires sur un Réseau – Modélisations, Statistique et Applications* (Masson, 1993).
24. P. Ishwar and P. Moulin, *On the equivalence of set-theoretic and maxent MAP estimation*, *IEEE Trans. Signal Process.,* **51** (2003) 698–713.
25. D. Leporini and J.-C. Pesquet, *Bayesian wavelet denoising: Besov priors and non-Gaussian noises, Signal Process.,* **81** (2001) 55–67.

26. S. Mallat, *A Wavelet Tour of Signal Processing,* 2nd ed. (Academic Press, 1999).
27. S. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, *IEEE Trans. Inform. Theory*, **32** (1992) 617–643.
28. Y. Meyer, *Ondelettes et Opérateurs – vol. I,* (Hermann, 1990).
29. S. Osher and N. Paragios (Eds.), *Geometric Level Set Methods in Imaging, Vision, and Graphics,* (Springer, 2003).
30. C. Sánchez-Avila, *A nonlinear adaptive wavelet-based method for spiky deconvolution, Nonlinear Anal.,* **47** (2001) 4937–4948.
31. C. Sánchez-Avila, *Wavelet domain signal deconvolution with singularity-preserving regularization, Math. Comput. Simul.,* **61** (2003) 165–176.
32. H. Stark (Ed.), *Image Recovery: Theory and Application* (Academic Press, 1987).
33. C. M. Stein, *Estimation of the mean of a multivariate normal distribution, Ann. Stat.*, **9** (1981) 1135–1151.
34. H. J. Trussell and M. R. Civanlar, *The feasible solution in signal restoration, IEEE Trans. Acoust., Speech, Signal Process.,* **32** (1984) 201–212.
35. H. J. Trussell, *A priori knowledge in algebraic reconstruction methods,* in *Advances in Computer Vision and Image Processing* (T. S. Huang, Ed.), **1** (JAI Press, 1984) 265–316.
36. J. Weickert, *Anisotropic Diffusion in Image Processing* (Teubner-Verlag, 1998).
37. D. C. Youla and H. Webb, *Image restoration by the method of convex projections: Part 1 – theory, IEEE Trans. Med. Imaging,* **1** (1982) 81–94.

Fig. 1. Original 256×256 Lena image.

Fig. 2. Noisy image – MSE $= 900$.

Fig. 3. Besov-constrained denoising with exact bounds – MSE = 179.

Fig. 4. Besov-constrained denoising with underestimated bounds – MSE = 268.

Fig. 5. Besov-constrained denoising with overestimated bounds – MSE = 700.

Fig. 6. Proposed wavelet-constrained denoising – MSE = 150.

Fig. 7. SUREshrink denoising – MSE = 202.

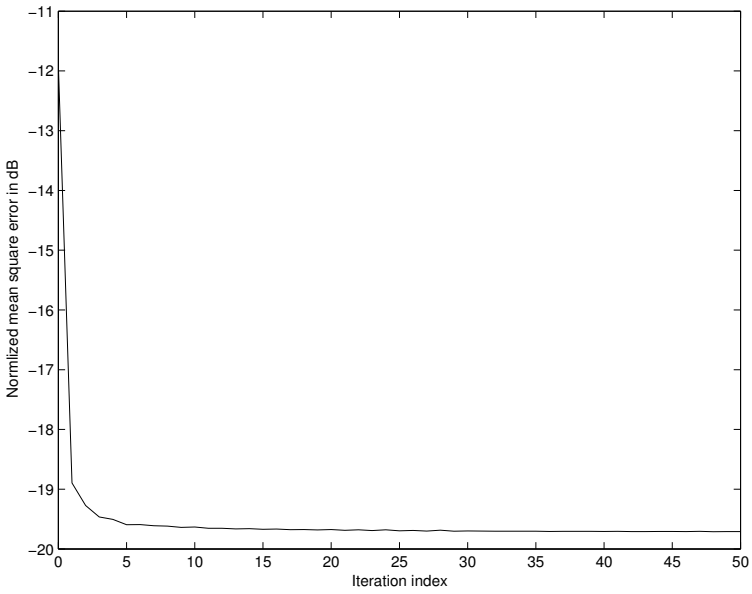

Fig. 8. $20\log_{10}(\|x_n - \overline{x}\|/\|\overline{x}\|)$ versus the iteration index n.