GROUP SPARSITY WITH OVERLAPPING PARTITION FUNCTIONS

Gabriel Peyré[1], *Jalal Fadili*[2]

[1]Ceremade
CNRS-Univ. Paris-Dauphine, France
Gabriel.Peyre@ceremade.dauphine.fr

[2]GREYC
CNRS-ENSICAEN-Univ. Caen, France
Jalal.Fadili@greyc.ensicaen.fr

ABSTRACT

This paper introduces a novel and versatile group sparsity prior for denoising and to regularize inverse problems. The sparsity is enforced through arbitrary block-localization operators, such as for instance smooth localized partition functions. The resulting blocks can have an arbitrary overlap, which is important to reduce visual artifacts thanks to the increased translation invariance of the prior. They are moreover not necessarily binary, and allow for non-integer block sizes. We develop two schemes, one primal and another primal-dual, originating from the non-smooth convex optimization realm, to efficiently solve a wide class of inverse problems regularized using this overlapping group sparsity prior. This scheme is flexible enough to handle both penalized and constrained versions of the optimization problems at hand. Numerical results on denoising and compressed sensing are reported and show the improvement brought by the overlap and the smooth partition functions with respect to classical group sparsity.

1. INTRODUCTION

Sparsity for denoising and inverse problems. Sparsity is a key concept used to solve various image processing problems. One of its earliest manifestations is through the seminal work by Donoho and Johnstone on thresholding operators in orthogonal bases for denoising [19]. Sparsity in orthogonal and redundant dictionaries, such as wavelets, has then been extensively used to attack a variety of inverse problems, by solving a wisely penalized least-squares problem, where the penalty is chosen to enforce the sparsity of the coefficients. In its simplest form, the prior penalty is the ℓ^1 norm of the coefficients. We refer to [27] for an overview of the methods and theoretical results pertaining to sparse regularization of inverse problems.

Non-overlapping group sparsity. It turns out that term-by-term sparsity is usually not enough to obtain state-of-the-art results both for denoising and inverse problems involving natural images. Indeed, wavelet coefficients of images are not only sparse, they typically exhibit local dependencies among neighboring coefficients. Geometric features (edges, textures) are poorly sparsified by isotropic multiscale decompositions and create such dependencies. Block thresholding operators group coefficients in non-overlapping blocks to take into account these dependencies and improve the performances both theoretically and in practice, see e.g. [5, 9, 30, 12] for denoising, and [13] for deconvolution. Convex block sparsity priors have also been used in machine learning, e.g. group-Lasso in [31, 1], as well as for inverse problems such as compressed sensing recovery, see [2] among others.

Overlapping group sparsity. To further improve the denoising performance, [6] proposed to take into account the energy of overlapping blocks to threshold non-overlapping groups of coefficients. Sparse group convex priors have been studied with overlapping blocks that have a chain structure in [29] or a tree structure in [25, 32]. These constrained structures lead to interesting properties of the patterns of non-zero coefficients and lead to efficient algorithms. Generic arbitrary overlapping blocks have been recently considered in [23] using a synthesis formulation (see (6) hereafter for more details). Similarly, [11] consider arbitrary blocks, but use an analysis formulation.

Convex optimization for sparse regularization. Sparse regularization requires solving challenging non-smooth convex optimization problems. The structure of variational problems with convex sparsity penalties, that are mostly variations around the ℓ^1 penalty, favors the use of proximal splitting schemes, see [15, 4] for review chapters. For instance, problems involving a smooth fidelity term and an ℓ^1 penalty, the one-step forward-backward (see e.g. [16, 17]), or its multi-step accelerated versions [3, 28] are excellent candidates. For overlapping block sparsity, one needs to use more sophisticated splitting schemes. Among them, we can think of the (primal) Douglas-Rachford (DR) algorithm [26, 14]), the dual scheme of the alternating direction method of multipliers [22], or primal-dual schemes such as [10, 8] and others. The choice of the minimization algorithm is conditioned by the structure of the objective functional at hand. In Section 3, we shall discuss two different algorithms.

Contributions. Our work is closely related to the analysis overlapping block sparsity prior [11], and extend it in several crucial aspects. (i) We introduce a novel generic prior for overlapping group sparsity, which can use smoothly overlapping partition functions. (ii) We develop two efficient algorithms (one primal and one primal-dual) that are flexible enough to solve exactly, and without any smoothing of the objective functional, a wide class of linear inverse optimization problems involving our group sparsity prior. (iii) We report a numerical study to outline the importance of group sparsity with smoothly overlapping partition functions for image denoising and linear inverse problem regularization.

2. SMOOTHLY OVERLAPPING GROUP SPARSITY

Inverse problem regularization. This paper aims at studying regularization schemes to solve the ill-posed linear in-

ISSN 2076-1465

verse problem that consists in recovering x_0 from $y = Ax_0 + w$ where $A : \mathbb{R}^N \to \mathbb{R}^P$ is a bounded linear operator, $x_0 \in \mathbb{R}^N$ is the (unknown) signal/image to recover, and $w \in \mathbb{R}^P$ is an additive noise. The classical regularization approach performs the recovery by solving

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - Ax\|^2 + \lambda J(x), \tag{1}$$

where $J(x)$ is some penalty functional that reflects the prior information about the signal to recover, and $\lambda > 0$ is a regularization parameter that should be adapted to the noise level. We assume in the sequel that $J(x)$ is a convex lower-semicontinuous (lsc) and proper function on $\mathbb{R}^N$, with $A(\mathrm{dom}(J)) \neq \emptyset$ and J is coercive if $\ker(A) \neq \{0\}$. The latter conditions ensure properness of the objective and existence of a minimizer. Note by the way that other data fidelity terms could be used instead of the quadratic term $\frac{1}{2}\|y - Ax\|^2$ to reflect some statistical knowledge about the noise w. We restrict our attention to this fidelity for simplicity of the exposition. The penalized (1) has an equivalent constrained form, in the sense that $\exists \varepsilon(\lambda) > 0$ such that the minimizer of (1) is also a solution to

$$\min_{x \in \mathbb{R}^N,\ \|y - Ax\| \leqslant \varepsilon(\lambda)} J(x)\ . \tag{2}$$

In the noiseless case, both formulations (1)-(2) reduce to

$$\min_{x \in \mathbb{R}^N,\ Ax = y} J(x)\ . \tag{3}$$

Group sparsity. We consider a family of priors that measures the sparsity of the signal x using a countable collection of localization operators $B_i : \mathbb{R}^N \to \mathbb{R}^{N_i}$ for $i \in I$. In practice, $B_i(x)$ depends only on a few values $x(t)$ of the signal. Thus, our group sparsity prior extends the classical ℓ^1-norm sparsity by considering

$$J(x) = \sum_{i \in I} \varphi_i(B_i x) = \Phi(Bx) \tag{4}$$

where each $\varphi_i : \mathbb{R}^{N_i} \to \mathbb{R}^+$ is a proper, lsc convex function, and we have used the shorthand notations

$$Bx = (B_i x)_{i \in I} \in \Omega = \prod_{i \in I} \mathbb{R}^{N_i}$$

$$\forall u = (u_i)_i \in \Omega, \quad \Phi(u) = \sum_{i \in I} \varphi_i(u_i).$$

It is sufficient to require that $\forall i$, φ_i is coercive and B is injective, and the objective is proper to ensure minimizer existence for (1)-(3).

Examples. A classical group sparsity regularization that have been considered in the literature uses the intra-block ℓ^p norm, for $p \geqslant 1$,

$$\varphi_i(v) = \left(\sum_{k=0}^{N_i - 1} |v(k)|^p \right)^{1/p},$$

with the classical modification for $p = +\infty$. Note that one should have $p > 1$ to promote block sparsity. The choice $p = 1$ yields the classical, non-grouped, ℓ^1 sparsity.

To perform block regularization, we propose to define the (diagonal) localization operators B_i through the partition functions $b_i(t) \geqslant 0$

$$B_i x = (b_i(t) x(t))_{t \in S_i} \tag{5}$$

where the support of $b_i(t)$ is $S_i = \{t \setminus b_i(t) \neq 0\}$ of size $|S_i| = N_i$. To regularize all the coefficients, we require that $\sum_i b_i(t) > 0$ for all t, and therefore $\ker(B) = \{0\}$. Note that the group-Lasso regularization is obtained by specializing b_i to binary functions $b_i(t) \in \{0, 1\}$, with non-overlapping supports and $\bigcup_{i \in I} S_i = \{0, \ldots, N-1\}$.

Analysis vs. synthesis for block sparsity. A family of overlapping block sparsity priors has been introduced in [23] for binary blocks. It extends trivially to our setting as follows

$$J^{\mathrm{synth}}(x) = \min_{u \in \Omega, B^* u = x} \Phi(u). \tag{6}$$

where B^* is the adjoint of the blocking operator B. Note that using this prior, the recovery problem (1) can be written as computing $x = B^* u$ where u solves

$$\min_{u \in \Omega} \frac{1}{2}\|y - AB^* u\|^2 + \lambda \Phi(u).$$

This corresponds to a "synthesis" formulation using a redundant dictionary Φ^*, as defined by [20], whereas (4) can be seen as an "analysis" formulation.

Note that since the blocks overlap, the analysis prior (4) and the synthesis prior (6) are expected to produce different results. It is however beyond the scope of this paper to provide a detailed analysis of the performances of these two classes of methods.

3. MINIMIZATION ALGORITHMS

All minimization problems considered in this paper can be written as

$$\min_{x \in \mathbb{R}^N} \Psi(Ax) + \lambda \Phi(Bx) \tag{7}$$

$$\text{where} \quad \forall g \in \mathbb{R}^P, \quad \Psi(g) = \begin{cases} \frac{1}{2\lambda}\|y - g\|^2 & \text{for} \quad (1), \\ i_{\|y - \cdot\| \leqslant \varepsilon}(g) & \text{for} \quad (2), \\ i_{y = \cdot}(g) & \text{for} \quad (3) \end{cases}$$

where i_C is the indicator of the closed convex set C, so that $i_C(g) = 0$ if $g \in C$ and $i_C(g) = +\infty$ otherwise. We of course suppose that $\exists x \in \mathbb{R}^N$ such that $Ax \in \mathrm{dom}(\Psi)$ and $Bx \in \mathrm{dom}(\Phi)$.

Sections (3.3) and (3.2) describe respectively a primal and a primal-dual approach to solve (7). It will turn out that the primal approach is quite efficient on the practical side, but is confined to problems where $A^* A$ can be diagonalized efficiently. The primal-dual does not have such a restriction. Having said that, it is beyond the scope of this paper to delve into a detailed comparison of convergence rates of existing methods to solve (7).

3.1 Proximity Operators

We assume that beside above assumptions on the φ_i's, these functions are also "almost" simple, meaning that the associated proximity operators can be computed either exactly in closed-form or approximately with a rapidly converging optimization scheme. Recall that the proximity operator (or proximal mapping) of a proper lsc and convex function φ_i is defined as

$$\forall\, v \in \mathbb{R}^{N_i}, \quad \operatorname{prox}_{\gamma\varphi_i}(v) = \operatorname*{argmin}_{w\in\mathbb{R}^{N_i}} \frac{1}{2}\|v-w\|^2 + \gamma\varphi_i(w),$$

see for instance the review papers [15, 4]. The proximity operator enjoys a whole calculus framework among which Moreau identity that will be useful in the sequel. Let φ_i^* be the Legendre-Fenchel conjugate of φ_i, then

$$\operatorname{prox}_{\gamma\varphi_i^*}(v) = v - \gamma \operatorname{prox}_{\varphi_i/\gamma}(v/\gamma)\ . \tag{8}$$

The ℓ^1 and ℓ^2 norms are simple functions, since

$$\begin{aligned}
\operatorname{prox}_{\gamma\|\cdot\|_1}(v) &= \left(\max\left(0, 1-\frac{\gamma}{v(k)}\right) v(k)\right)_k,\\
\operatorname{prox}_{\gamma\|\cdot\|_2}(v) &= \max\left(0, 1-\frac{\gamma}{\|v\|_2}\right) v.
\end{aligned}$$

For the other ℓ^p norms (as well as for any 1-homogeneous function), it can be computed by invoking conjugacy arguments and Moreau Identity (8), which yields

$$\operatorname{prox}_{\gamma\|\cdot\|_p}(v) = v - \gamma \operatorname{Proj}_{\|\cdot\|_q \leqslant 1}(v/\gamma)$$

where $\operatorname{Proj}_{\|\cdot\|_q\leqslant 1}$ is the orthogonal projector on the dual ℓ^q ball (i.e. $1/p+1/q=1$). For $p=\infty$, this can be computed using the projection on the ℓ^1 ball (see for instance [21]), and for the other values, the projector on the ℓ^q ball can be computed using a few Newton iterations [24].

Owing to separability of Φ, its proximal mapping is simply the concatenation of those of the φ_i's,

$$\forall\, u=(u_i)_{i\in I} \in \Omega, \quad \operatorname{prox}_{\gamma\Phi}(u) = \left(\operatorname{prox}_{\gamma\varphi_i}(u_i)\right)_{i\in I}. \tag{9}$$

3.2 Primal Algorithm

In order to apply the DR splitting scheme [26] (see also [14] and the review paper [15]), we re-write the optimization problem (7) by introducing an auxiliary variable $u\in\Omega$ and the linear constraint $u = B(x)$

$$\min_{z=(x,u)\in\mathbb{R}^N\times\Omega} H(z) + i_{\mathscr{C}}(z)$$

where $H(x,u) = \Psi(A(x)) + \lambda\Phi(u)$. Note that in this section, $z=(x,u)\in\mathbb{R}^N\times\Omega$ actually denotes a couple of variables so we write indifferently $H(z)$ and $H(x,u)$. The linear constraint is defined by $\mathscr{C} = \{(x,u) \setminus u = Bx\}$, i.e. $z \in \ker([-B \ \mathrm{Id}])$.

The proximal mapping of H is easily accessible as

$$\operatorname{prox}_{\gamma H}(x,u) = \left(\operatorname{prox}_{\gamma\Psi\circ A}(x), \operatorname{prox}_{\gamma\lambda\Phi}(u)\right)$$

where

$$\operatorname{prox}_{\gamma\Psi\circ A}(x) = \begin{cases} \left(\mathrm{Id}_N + \frac{\gamma}{\lambda}A^*A\right)^{-1}\left(x+\frac{\gamma}{\lambda}A^*y\right) & \text{for} \quad (1),\\ x + A^*(AA^*)^{-1}(y-Ax) & \text{for} \quad (3),\end{cases}$$

and the proximity operator of Φ is defined in (9). It is thus possible to compute efficiently this proximity operator if A^*A can be efficiently diagonalized. Note also that there is no closed-form expression of $\operatorname{prox}_{\Psi\circ A}$ in the case (2) unless A is a tight frame.

The orthogonal projector onto $\mathscr{C}$ is such that

$$\operatorname{prox}_{i_{\mathscr{C}}}(x,u) = \operatorname{Proj}_{\mathscr{C}}(x,u) = (\tilde{x}, B\tilde{x})$$

where $\tilde{x} = (\mathrm{Id}_N + B^*B)^{-1}(B^*u+x)$. In the specific case of a diagonal block operator B of the form (5), this projection can be computed in $O(N)$ operation since the operator B^*B is diagonal. Indeed,

$$B^*B = \operatorname{diag}\Big(\sum_{t\in S_i} b_i(t)^2\Big)_t.$$

Given some $\tilde{z}^{(0)} = (x^{(0)}, u^{(0)})$, the DR algorithm is then summarized as follows:

$$\begin{aligned}
\tilde{z}^{(n+1)} &= \left(1-\frac{\mu}{2}\right)\tilde{z}^{(n)} + \frac{\mu}{2}\mathrm{rPox}_{i_{\mathscr{C}}}(\mathrm{rProx}_{\gamma H}(\tilde{z}^{(n)}))\\
z^{(n+1)} &= \mathrm{Prox}_{\gamma H}(\tilde{z}^{(n+1)})
\end{aligned}$$

for $\mu\in]0,2[$ and $\gamma>0$, where have use the following shorthand notation

$$\mathrm{rProx}_{\gamma H}(z) = 2\mathrm{Prox}_{\gamma H}(z) - z,$$

see for instance [14]. It can be shown that the sequence $z^{(n)} \to z^\star = (x^\star, u^\star)$ as $n\to+\infty$, where $x^\star$ is a (global) minimizer of (7).

3.3 Primal Dual Algorithm

Both the penalized (1) and constrained (2)-(3) problems can be cast as the minimization of $F(Kx)$ where $\forall (g,u)\in \mathbb{R}^P\times\Omega$,

$$F(g,u) = \Psi(g) + \lambda\Phi(u) \quad \text{and} \quad Kx = (Ax, Bx) \in \mathbb{R}^P\times\Omega.$$

By separability, we have $\operatorname{prox}_{\gamma F}(g,u) = (\tilde{g}, \operatorname{prox}_{\gamma\Phi}(u))$ with

$$\tilde{g} = \begin{cases} \frac{g+\lambda\gamma y}{1+\lambda\gamma} & \text{for} \quad (1),\\ y + \varepsilon\frac{u-y}{\max(\varepsilon,\|u-y\|)} & \text{for} \quad (2),\\ y & \text{for} \quad (3).\end{cases}$$

Since F has an explicit proximal operator and K is a bounded linear operator, our functionals can be minimized efficiently using a primal-dual scheme such as the one proposed in [8] (anoother potential candidate is e.g. [10]).

Given $x^{(0)} = \bar{x}^{(0)} \in \mathbb{R}^N$ and $\beta^{(0)}\in\mathbb{R}^P\times\Omega$, define the sequence of iterates:

$$\begin{aligned}
\beta^{(n+1)} &= \operatorname{prox}_{\sigma F^*}\left(\beta^{(n)} + \sigma K\bar{x}^{(n)}\right)\\
x^{(n+1)} &= x^{(n)} - \tau K^*\beta^{(n+1)}\\
\bar{x}^{(n+1)} &= x^{(n+1)} + \theta(x^{(n+1)} - x^{(n)})\ .
\end{aligned}$$

With the proviso that $0<\theta\leqslant 1$ and $\tau\sigma\|K\|^2<1$. For $\theta=1$, it is shown in [8] that $x^{(n)}\to x^\star$ as $n\to+\infty$, where $x^\star$ is a (global) minimizer of $F(Kx)$.

4. NUMERICAL ILLUSTRATIONS

In the numerical examples, we consider a natural image f_0 of $N = n \times n$ pixels, where $n = 256$, that is normalized so that $\|f_0\|_\infty = 1$. We implement our method using overlapping smooth partition functions as defined in (5). For simplicity we use a translation invariant collection of 2-D Gaussian partition functions, where each $b_i(t)$ is centered around pixel $i = (i_1, i_2)$ and has a variance s^2

$$\forall t = (t_1, t_2) \in \{0, \dots, n-1\}^2, \quad b_i(t) = e^{-\frac{\|t-i\|^2}{s^2}} 1_{S_i}(t), \quad (10)$$

where 1_{S_i} restricts the support of b_i to

$$S_i = \{[i_1 - 3s], \dots [i_1 + 3s]\} \times \{[i_2 - 3s], \dots [i_2 + 3s]\}$$

where $[\cdot]$ is the nearest integer rounding operator.

We bench the efficiency of our approach for a varying value of $s > 0$, which parameterizes the effective width of the overlapping partition functions. We also compare these results with the classical block sparsity prior without overlapping, which corresponds to using $b_i(t) = 1_{S_i}(t)$ where the S_i's are blocks of size $w \times w$

$$S_i = \{i_1 w, \dots (i_1 + 1)w - 1\} \times \{i_2 w, \dots (i_2 + 1)w - 1\}. \quad (11)$$

4.1 Denoising

In the denoising experiment, we observe a noisy image $f_0 + \tilde{w}$ where $f_0 \in \mathbb{R}^N$ is the (unknown) clean image and $\tilde{w} \sim \mathcal{N}(0, \sigma^2)$. We use a bi-orthogonal 7-9 wavelet transform W to compute the coefficients $y = W(f_0 + \tilde{w}) = x_0 + w$ where $x_0 = W(f_0)$ are the (unknown) coefficients to estimate and w remains $\mathcal{N}(0, \sigma^2)$. The denoised coefficients x are estimated by solving (1), and the denoised image is recovered as $f = W^{-1}(x)$. We tested different values of $\lambda > 0$ so as to maximize the $\mathrm{PSNR}(f_0, f) = -10\log_{10}(\|f - f_0\|^2/N)$.

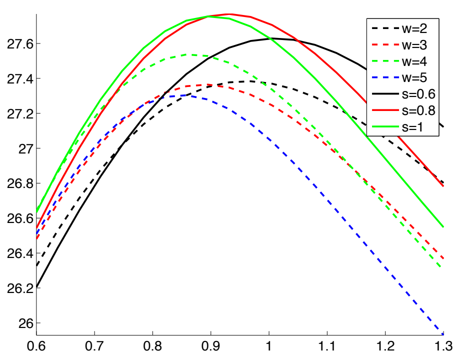

Figure 1: *Evolution of the PSNR as a function of λ/σ, for sparse regularization with non-overlaping blocks (dashed lines, the different curves correspond to different block sizes w) and overlapping smooth partition functions (solid lines, the different curves correspond to different widths s of the Gaussian).*

Figure 1 shows the evolution of the PSNR with the regularization parameter λ, for the boat image, displayed on Figure 2. The best block size for non-overlapping regularization is $w = 4$. The best partition width s for overlapping regularization is $s = 0.8$. The PSNR gain is about 0.3dB, which is a modest improvement, but is consistent across a wide range of natural images.

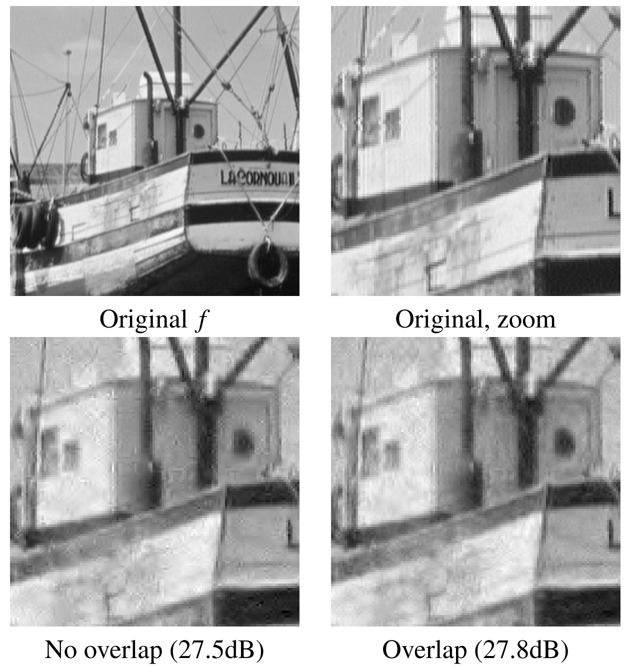

Figure 2: *Comparison of denoising using non-overlapping blocks* (11) *and overlapping partition functions* (10).

4.2 Compressed Sensing

We consider a noiseless compressed sensing recovery problem, where a small number $P < N$ of measurements $y = Mf_0$ are collected with a linear sensing operator $M \in \mathbb{R}^{P \times N}$, which is a realization from a random matrix ensemble. This corresponds to a stylized and idealized compressed sensing acquisition scenario, as proposed by Candès, Romberg and Tao and [7] and Donoho [18] to jointly sample and compress sparse signals. We consider here $P = N/8$, and $Mf = (P_2 \circ D \circ P_1(f)) \downarrow_P$, where P_1, P_2 are realizations of random permutations of $\{0, \dots, N-1\}$, D is an orthogonal discrete cosine transform, and $\downarrow_P$ selects the P first entries of a vector. This operator M is both random (thus enabling a provably efficient recovery from a small number of measurements), and can be computed in $O(N \log(N))$ operations. The coefficients x of the recovered signal $f = W^{-1}(x)$ are obtained by solving the constrained formulation (3) with $A = MW^{-1}$.

Figure 3 shows the evolution of the PSNR as a function of the partition function width s. Its optimal value on this example is $s = 1$. The gain with respect to non-overlapping blocks of optimal size $w = 4$ is roughly $0.45dB$. Figure 4 shows a visual comparison of the two priors.

Conclusion

This paper has introduced a novel group sparsity prior to regularize inverse problems together with accompanying optimization algorithms. This allows to use structured group sparsity with smoothly overlapping partition functions. We believe that this prior is a serious option to consider and will

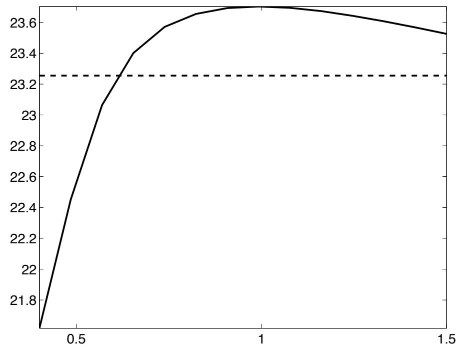

Figure 3: *Evolution of the PSNR as a function of of the partition function width s defined in* (10)*. As a reference for comparison, the dashed line corresponds to the PSNR obtained with non-overlapping blocks* (11) *using* $w = 4$.

No overlap (23.25dB) Overlap (23.7dB)

Figure 4: *Comparison of compressed sensing recovery using non-overlapping blocks* (11) *and overlapping smooth partition functions* (10).

be useful for a variety of applications, for instance when binary blocks are too constrained, and translation invariance of the regularization is a desirable property. Numerical experiments on natural images show that this brings some improvement with respect to classical block sparsity using non-overlapping blocks.

REFERENCES

[1] F. R. Bach. Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.*, 9:1179–1225, June 2008.

[2] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inf. Theo.*, 56:1982–2001, 2010.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *Journal on Imaging Sciences*, 2(1):183–202, 2009.

[4] A. Beck and M. Teboulle. *Gradient-Based Algorithms with Applications to Signal Recovery Problems*, chapter Convex Optimization in Signal Processing and Communications, ed. Y. Eldar and D. Palomar. Cambridge university press, 2010.

[5] T. Cai. Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist*, 27(3):898–924, 1999.

[6] T. T. Cai and B. W. Silverman. Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya*, 63:127–148, 2001.

[7] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory*, 52(2):489–509, 2006.

[8] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Technical report, preprint CMAP-685, 2010.

[9] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A nonlinear stein based estimator for multichannel image denoising. *IEEE Trans. on Sig.*, 56(8):3855–3870, 2008.

[10] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Math. Program.*, 64:81–101, 1994.

[11] X. Chen, Q. Lin, S. Kim, J. G. Carbonell, and E. P. Xing. An efficient proximal-gradient method for general structured sparse learning. *Preprint arXiv:1005.4717*, 2010.

[12] C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for image denoising. *App. Comp. Harm. Anal.*, 28:67–88, 2009.

[13] C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for wavelet-based image deconvolution. *Elec. J. Statistics*, 4:415–435, 2010.

[14] P. L. Combettes and J.-C. Pesquet. A douglas-rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.

[15] P. L. Combettes and J.-C. Pesquet. *Proximal splitting methods in signal processing*, chapter Fixed-Point Algorithms for Inverse Problems in Science and Engineering, ed. H. H. Bauschke et al., pages 185–212. Springer-Verlag, 2011.

[16] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4), 2005.

[17] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. on Pure and Appl. Math.*, 57:1413–1541, 2004.

[18] D. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, 2006.

[19] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[20] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947, 2007.

[21] J. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, 20(3):657–669, 2011.

[22] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Computational Mathematics with Applications*, 2:17?40, 1976.

[23] L. Jacob, G. Obozinski, and J-P. Vert. Group lasso with overlap and graph lasso. In Andrea Pohoreckyj Danyluk, Léon Bottou, and Michael L. Littman, editors, *Proc. ICML 2009*, volume 382, page 55, 2009.

[24] L. Jacques, D. Hammond, and M.J. Fadili. Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine. *IEEE Transactions on Information Theory*, 57(1):559–571, 2011.

[25] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *ICML*, 2010.

[26] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[27] S. Mallat. *A Wavelet Tour of Signal Processing, 3rd edition*. Elsevier, 2009.

[28] Y. Nesterov. Gradient methods for minimizing composite objective function. CORE Discussion Papers 2007076, UCL, September 2007.

[29] R. Tibshirani and M. Saunders. Sparsity and smoothness via the fused lasso. *Journal of the Royal Stat. Soc., Series B*, 61(1):91–108, 2005.

[30] G. Yu, S. Mallat, and E. Bacry. Audio denoising by time-frequency block thresholding. *IEEE Trans. on Sig.*, 56(5):1830–1839, 2008.

[31] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. of The Roy. Stat. Soc. B*, 68(1):49–67, 2006.

[32] P. Zhao, G. Rocha, and B. Yu. Grouped and hierarchical model selection through composite absolute penalties. *Annals of Statistics*, 37(6A):3468–3497, March 2009.