A proximal approach for constrained cosparse modelling

Giovanni Chierchia, Nelly Pustelnik, Jean-Christophe Pesquet, Béatrice Pesquet-Popescu

A PROXIMAL APPROACH FOR CONSTRAINED COSPARSE MODELLING

G. Chierchia,[1] *N. Pustelnik,*[2] *J.-C. Pesquet,*[3] *and B. Pesquet-Popescu,*[1]

[1] Télécom ParisTech/Institut Télécom, LTCI, UMR CNRS 5141, 75014 Paris, France

[2] Laboratoire de Physique de l'ENS Lyon, UMR CNRS 5672, F69007 Lyon, France

[3]Université Paris-Est, Lab. d'Informatique Gaspard Monge, UMR CNRS 8049, 77454 Marne la Vallée, France

ABSTRACT

The concept of cosparsity has been recently introduced in the arena of compressed sensing. In cosparse modelling, the ℓ_0 (or ℓ_1) cost of an analysis-based representation of the target signal is minimized under a data fidelity constraint. By taking benefit from recent advances in proximal algorithms, we show that it is possible to efficiently address a more general framework where a convex block sparsity measure is minimized under various convex constraints. The main contribution of this work is the introduction of a new epigraphical projection technique, which allows us to consider more flexible data fidelity constraints than the standard linear or quadratic ones. The validity of our approach is illustrated through an application to an image reconstruction problem in the presence of Poisson noise.

***Index Terms*—** Signal restoration, compressed sensing, iterative methods, optimization methods, wavelet transforms.

1. INTRODUCTION

The tremendous research activity in the field of sparse modelling and compressed sensing [1, 2] has been at the root of novel signal/image processing concepts. Among the most appealing ideas, the concept of cosparsity has been introduced [3] to extend the standard approach where the ℓ_0 cost (or its ℓ_1 convex relaxation) of the target signal is minimized under a linear constraint accounting for the observation model. In cosparse modelling, a sparsity measure of a linear transformation Fx of the sought signal x is minimized subject to a linear constraint. In the presence of noise this constraint is usually replaced by a quadratic one of the form

$$\|Tx - z\|^2 \leq \eta \tag{1}$$

where z is the vector of observations, T is the linear operator modelling the observation process and η is a positive bound related to the variance of the noise. Instead of considering a standard ℓ_0 or ℓ_1 cost function, it may be fruitful to consider more general block or group sparsity measures [4, 5, 6, 7] for which appropriate algorithmic tools can also be developed.

For a long time in signal and image processing, it has been recognized that incorporating as many available information as possible on the sought solution to an inverse problem is beneficial to the estimation performance. POCS algorithm [8] and its extensions have been developed to iteratively compute the solution to a convex admissibility problem, where the target signal is known to belong to the intersection of a finite number of convex constraint sets. Recent offsprings of these methods and their applications in learning are discussed in [9].

This work was supported by the Agence Nationale de la Recherche under grant ANR-09-EMER-004-03.

A natural extension of the notion of projection is provided by the proximity operator [10]. This tool constitutes the cornerstone of many convex optimization algorithms offering the ability to tackle nonsmooth functions (e.g. convex sparsity measures) and to split the optimized criterion in a sum of simpler terms onto which parallel processing can be performed.

Our objective in this paper is to provide efficient solutions to a convex formulation of cosparse modelling problems by taking benefit from recent advances in proximal algorithms. One of the main advantages of our proposed approach is that it allows us to address more flexible data fidelity constraints than (1) as well as additional convex constraints modelling prior information. It is worth emphasizing that the proposed method does not require to formulate the problem through its Lagrange dual, which would necessitate to determine the associated Lagrange multipliers. The flexibility of our approach is obtained by introducing a new epigraphical projection technique which is the main contribution of this work.

The paper is organized as follows. In Section 2, we formulate the considered problem. We motivate the choice of proximal tools in Section 3. In order to efficiently deal with a general data fidelity constraint, a splitting approach involving an epigraphical projection is proposed. This projection is described in Section 4 and its application is the topic of Section 5. Experiments in the context of an image reconstruction problem where the data are corrupted with a Poisson noise are presented in Section 6. Finally, some conclusions are drawn in Section 7.

2. PROBLEM STATEMENT

Let $\mathcal{H}$ be a real Hilbert signal space. We will denote by $\Gamma_0(\mathcal{H})$ the class of lower-semicontinuous functions from $\mathcal{H}$ to $]-\infty,+\infty]$. For simplicity, the norm of a Hilbert space is denoted by $\|\cdot\|$ without specifying the underlying space. Let F be a bounded linear operator from $\mathcal{H}$ to $\ell^2(\mathbb{K})$ with $\mathbb{K} \subset \mathbb{N}$,[1] which corresponds to some analysis transform, e.g. the decomposition onto some analysis frame. Let T be a linear operator from $\mathcal{H}$ to $\mathbb{R}^K$ modelling an observation process and let z be the associated observed vector in $\mathbb{R}^K$. We will further assume that the observation model yields an inequality of the form

$$g(T\overline{x}, z) \leq \eta \tag{2}$$

where $\overline{x} \in \mathcal{H}$ is the signal to be determined, $g(\cdot, z) \in \Gamma_0(\mathbb{R}^K)$ and $\eta \in \mathbb{R}$. For example, when the observation is generated through a linear additive noise model with log-concave noise probability density function $\propto \exp\big(-\Psi(\cdot)\big)$, we can choose $g(\cdot, z) = \Psi(\cdot - z)$ and the upper bound η can be fixed based on some given confidence level. We will also consider a nonempty closed convex set $C \subset \mathcal{H}$

[1] $\ell^2(\mathbb{K})$ denotes the space of square summable sequences indexed on $\mathbb{K}$.

modelling some prior knowledge (e.g. range value constraints) one may have on the sought signal $\overline{x}$.

The objective of this paper is to find an estimate of $\overline{x}$ by solving the following optimization problem

$$\underset{x}{\text{minimize}} \;\; \|Fx\|_{2,1} \quad \text{subject to} \quad \begin{cases} x \in C \\ g(Tx, z) \leq \eta. \end{cases} \tag{3}$$

Hereabove, $\|\cdot\|_{2,1}$ corresponds to a block (or group) sparsity measure taking the form

$$\big(\forall y \in \ell^2(\mathbb{K})\big) \qquad \|y\|_{2,1} = \sum_{b \in \mathbb{L}} \|B_b y\| \tag{4}$$

where $\mathbb{L} \subset \mathbb{K}$ and, for every $b \in \mathbb{L}$, B_b is some *block selection transform*. A linear transform B from $\ell^2(\mathbb{K})$ to $\mathbb{R}^L$ will be said to be a block selection transform if it allows us to select a block of L data from its input vector. This means that there exist distinct indices $k_1, \ldots, k_L$ in $\mathbb{K}$ such that, for every $y = (y^{(k)})_{k \in \mathbb{K}} \in \ell^2(\mathbb{K})$, $By = (y^{(k_1)}, \ldots, y^{(k_L)})$. Note that the operators $(B_b)_{b \in \mathbb{L}}$ may correspond to overlapping blocks having different sizes. However, for computational reasons, it will be assumed that these blocks can be regrouped into S sets of non-overlapping blocks. (This assumption is always satisfied when $\mathbb{L}$ is finite.) In other words, we assume that there exists a partition of $\mathbb{L}$ in S subsets $(\mathbb{L}_s)_{1 \leq s \leq S}$ such that

$$\big(\forall y \in \ell^2(\mathbb{K})\big) \quad \|y\|_{2,1} = \sum_{s=1}^{S} f_s(y) \tag{5}$$

where, for every $s \in \{1, \ldots, S\}$, the function f_s is defined as

$$\big(\forall y \in \ell^2(\mathbb{K})\big) \qquad f_s(y) = \sum_{b \in \mathbb{L}_s} \|B_b y\| \tag{6}$$

and $(B_b)_{b \in \mathbb{L}_s}$ select disjoint blocks. In the particular case when $S = 1$, $\mathbb{L} = \mathbb{L}_1 = \mathbb{K}$ and, for every $b \in \mathbb{L}$, $B_b \colon (y^{(k)})_{k \in \mathbb{K}} \mapsto y^{(b)}$, the classical ℓ^1-norm is obtained.

It is then useful to note that the considered minimization problem (3) can be reformulated more concisely as follows:

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \sum_{s=1}^{S} f_s(Fx) + \iota_C(x) + \iota_D(Tx) \tag{7}$$

where ι_C is the indicator function of C, which is equal to 0 on C and $+\infty$ on $\mathcal{H} \setminus C$. ι_D is the indicator function of the closed convex set D corresponding to the lower level set of $g(\cdot, z)$ at height η:

$$D = \big\{u \in \mathbb{R}^K \;\big|\; g(u, z) \leq \eta\big\} = \operatorname{lev}_{\leq \eta} g(\cdot, z). \tag{8}$$

3. PROXIMAL TOOLS

Problem (7) corresponds to the minimization of a sum of $S+2$ lower-semicontinuous convex functions composed with linear operators. It can thus be efficiently solved by proximal methods [10] (see also [11] for related augmented Lagrangian techniques). If we do not impose specific assumptions on the operators F and T, primal-dual proximal approaches [12, 13, 14, 15] appear as appealing methods to solve the problem. These methods have become popular recently due to the fact that they do not require operator inversions whose computational cost may be high. In order to apply these methods, we need however to compute the proximity operators of the involved functions in Problem (7).

Recall that the proximity operator of a function φ in $\Gamma_0(\mathcal{H})$ is defined as

$$\operatorname{prox}_{\varphi} \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{p \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{2}\|p - x\|^2 + \varphi(p). \tag{9}$$

The computation of the proximity operators of block sparsity measures has been investigated in [6]. Based on basic properties of proximity operators, it can be shown that, for every $s \in \{1, \ldots, S\}$ and for every $y \in \ell^2(\mathbb{K})$, $p = \operatorname{prox}_{\gamma f_s} y$ with $\gamma > 0$ is given by the following block thresholding rule:

$$(\forall b \in \mathbb{L}_s) \;\; B_b p = \begin{cases} \Big(1 - \dfrac{\gamma}{\|B_b y\|}\Big) B_b y & \text{if } \|B_b y\| > \gamma \\ (0, \ldots, 0) & \text{otherwise.} \end{cases} \tag{10}$$

The proximity of the indicator function of C reduces to the projection P_C onto this closed convex set. In the following, it will be assumed that the projection onto C has a closed form expression.

In turn, we will be interested in the case when g does not necessarily takes a simple form (e.g. a quadratic one), so that the projection onto D cannot be calculated in an explicit manner. In the next section, we describe a method allowing us to circumvent this difficulty.

4. EPIGRAPHICAL PROJECTION

Let us now see how to handle a convex constraint $\widetilde{D}$ of the form

$$\widetilde{D} = \big\{v \in \mathbb{R}^{KM} \;\big|\; h(v) \leq \eta\big\} \tag{11}$$

when h is a function in $\Gamma_0(\mathbb{R}^{KM})$ having the following separable form:

$$(\forall v \in \mathbb{R}^{KM}) \qquad h(v) = \sum_{r=1}^{K} h_r(\mathsf{v}^{(r)}). \tag{12}$$

Hereabove, the generic vector v has been decomposed into K blocks of coordinates as follows

$$v^\top = [\underbrace{(\mathsf{v}^{(1)})^\top}_{\text{size } M}, \ldots, \underbrace{(\mathsf{v}^{(K)})^\top}_{\text{size } M}] \tag{13}$$

and, for every $r \in \{1, \ldots, K\}$, $\mathsf{v}^{(r)} \in \mathbb{R}^M$ and h_r is a function in $\Gamma_0(\mathbb{R}^M)$.

By defining an auxiliary vector $\zeta \in \mathbb{R}^K$, the components of which are denoted by $\big(\zeta^{(r)}\big)_{1 \leq r \leq K}$. The inequality in (11) can be equivalently rewritten as

$$\sum_{r=1}^{K} \zeta^{(r)} \leq \eta \tag{14}$$

$$(\forall r \in \{1, \ldots, K\}) \qquad h_r(\mathsf{v}^{(r)}) \leq \zeta^{(r)}. \tag{15}$$

The latter inequality means that, for every $r \in \{1, \ldots, K\}$, $(\mathsf{v}^{(r)}, \zeta^{(r)})$ belongs to the so-called epigraph $\operatorname{epi} h_r$ of h_r. Let us now introduce the closed half-space of $\mathbb{R}^K$ defined as

$$V = \big\{\zeta \in \mathbb{R}^K \;\big|\; 1_K^\top \zeta \leq \eta\big\}, \quad \text{with } 1_K = (1, \ldots, 1)^\top \in \mathbb{R}^K, \tag{16}$$

and the closed convex set

$$E = \{(v, \zeta) \in \mathbb{R}^{KM} \times \mathbb{R}^K \mid (\forall r \in \{1, \ldots, K\}) \; (\mathsf{v}^{(r)}, \zeta^{(r)}) \in \operatorname{epi} h_r\}. \tag{17}$$

Then, Constraint (14) means that $\zeta \in V$, whereas Constraint (15) is equivalent to $(v, \zeta) \in E$. In other words, the constraint set $\widetilde{D}$ can be split into the two constraint sets V and E provided that K additional scalar variables $(\zeta^{(r)})_{1\leq r\leq K}$ are introduced. Dealing with extra constraints in the original problem is not a problem for the proximal algorithms we employ, as far as the projections onto the associated constraint sets can be easily computed.

In the present case, the projection onto V is simply given by

$$(\forall \zeta \in \mathbb{R}^K) \qquad P_V(\zeta) = \begin{cases} \zeta & \text{if } \mu \geq 0 \\ \zeta + \mu 1_K & \text{otherwise,} \end{cases} \tag{18}$$

where $\mu = \frac{1}{K}\big(\eta - 1_K^\top \zeta\big)$. On the other hand, the projection onto E is given by

$$(\forall (v, \zeta) \in \mathbb{R}^{KM} \times \mathbb{R}^K) \qquad P_E(v, \zeta) = (p, \theta) \tag{19}$$

where $\theta = (\theta^{(1)}, \dots \theta^{(K)})^\top$, vector $p \in \mathbb{R}^{KM}$ is blockwise decomposed as $p^\top = \big((\mathsf{p}^{(1)})^\top, \dots, (\mathsf{p}^{(K)})^\top\big)$ similarly to (13), and

$$(\forall r \in \{1, \dots, K\}) \quad (\mathsf{p}^{(r)}, \theta^{(r)}) = P_{\operatorname{epi} h_r}(\mathsf{v}^{(r)}, \zeta^{(r)}). \tag{20}$$

So, the problem reduces to the lower-dimensional problem of the determination of the projection onto the convex subset $\operatorname{epi} h_r$ of $\mathbb{R}^M \times \mathbb{R}$, for each $r \in \{1, \dots, K\}$. These projections have a closed form expression in a number of cases of interest an example of which is given below.

Proposition 4.1 *Assume that*

$$\begin{aligned} &\big(\forall \mathsf{v}^{(r)} = (\mathsf{v}^{(r,1)}, \dots, \mathsf{v}^{(r,M)})^\top \in \mathbb{R}^M\big) \\ &\qquad h_r(\mathsf{v}^{(r)}) = \max\{\mathsf{v}^{(r,j)} + \eta^{(r,j)} \mid 1 \leq j \leq M\} \end{aligned} \tag{21}$$

where $r \in \{1, \dots, R\}$ and $(\eta^{(r,1)}, \dots, \eta^{(r,M)})^\top \in \mathbb{R}^M$. For every $\mathsf{v}^{(r)} \in \mathbb{R}^M$, let $(\nu^{(r,j)})_{1\leq j\leq M}$ be a sequence obtained by sorting $(\mathsf{v}^{(r,j)} + \eta^{(r,j)})_{1\leq j\leq M}$ in ascending order, and set $\nu^{(r,0)} = -\infty$ and $\nu^{(r,M+1)} = +\infty$. Then, for every $\zeta^{(r)} \in \mathbb{R}$, $(\mathsf{p}^{(r)}, \theta^{(r)}) = P_{\operatorname{epi} h_r}(\mathsf{v}^{(r)}, \zeta^{(r)})$ is such that $\mathsf{p}^{(r)} = (\mathsf{p}^{(r,1)}, \dots, \mathsf{p}^{(r,M)})^\top$ where, for every $j \in \{1, \dots, M\}$,

$$\mathsf{p}^{(r,j)} = \begin{cases} \mathsf{v}^{(r,j)} & \text{if } \mathsf{v}^{(r,j)} \leq \theta^{(r)} - \eta^{(r,j)} \\ \theta^{(r)} - \eta^{(r,j)} & \text{otherwise,} \end{cases} \tag{22}$$

$\theta^{(r)} = \frac{1}{M-\overline{\jmath}+2}\Big(\zeta^{(r)} + \sum_{j=\overline{\jmath}}^{M} \nu^{(r,j)}\Big)$ and $\overline{\jmath}$ is the unique integer in $\{1, \dots, M+1\}$ such that

$$(M - \overline{\jmath} + 2)\nu^{(r,\overline{\jmath}-1)} < \zeta^{(r)} + \sum_{j=\overline{\jmath}}^{M} \nu^{(r,j)} \leq (M - \overline{\jmath} + 2)\nu^{(r,\overline{\jmath})}.$$

As will be shown in the next section, this result will play a central role in the solution of Problem (7) for non-trivial forms of function g.

5. ALGORITHMIC SOLUTION

Let us come back to the original problem formulated in Section 2 in the case when, for every $u = (u^{(1)}, \dots, u^{(K)})^\top \in \mathbb{R}^K$,

$$g(u, z) = \sum_{r=1}^{K} g_r(u^{(r)}, z^{(r)}). \tag{23}$$

In the above expression, $(z^{(r)})_{1\leq r\leq K}$ are the components of z, and, for every $r \in \{1, \dots, K\}$, $g_r(\cdot, z^{(r)})$ is a one-variable function in $\Gamma_0(\mathbb{R})$. For example, such a separable function g may correspond to the negative-log-likelihood of an independent noise model. Let $r \in \{1, \dots, K\}$ and let $a_1^{(r)}, \dots, a_M^{(r)}$ be given distinct elements of the domain $g_r(\cdot, z^{(r)})$. For every $u^{(r)} \in \mathbb{R}$, the following inequality holds:

$$g_r(u^{(r)}, z^{(r)}) \geq g_r(a_j^{(r)}, z^{(r)}) + \delta_j^{(r)}(u^{(r)} - a_j^{(r)}) \tag{24}$$

where, for every $j \in \{1, \dots, M\}$, $\delta_j^{(r)} \in \mathbb{R}$ is any subgradient of $g_r(\cdot, z_r)$ at $a_j^{(r)}$. A lower approximation of $g_r(\cdot, z^{(r)})$ is then given by $h_r \circ \Delta^{(r)}$ where $\Delta^{(r)} = [\delta_1^{(r)}, \dots, \delta_M^{(r)}]^\top$ and h_r satisfies (21) with

$$\eta^{(r,j)} = g_r(a_j^{(r)}, z^{(r)}) - \delta_j^{(r)} a_j^{(r)}. \tag{25}$$

By defining

$$\Delta = \begin{bmatrix} \Delta^{(1)} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Delta^{(K)} \end{bmatrix} \in \mathbb{R}^{KM \times K} \tag{26}$$

we have then

$$g(u, z) \simeq h(\Delta u) \tag{27}$$

where the approximation can be as close as desired by choosing M large enough.

Using the notations of Section 4, Problem (7) can then be recast as the following optimization problem in the product space $\mathcal{H} \times \mathbb{R}^K$:

$$\underset{(x,\zeta)\in\mathcal{H}\times\mathbb{R}^K}{\text{minimize}} \;\; \iota_C(x) + \iota_V(\zeta) + \iota_E(\Delta T x, \zeta) + \sum_{s=1}^{S} f_s(Fx). \tag{28}$$

The projection onto V is given by (18), the projection onto E by Proposition 4.1 and the proximity operator of the functions f_s by (10). In addition, the projection onto C has been assumed to take an explicit form. As mentioned in Section 3, various proximal methods can thus be applied to solve this nonsmooth optimization problem. In this work, we will employ the Monotone+Lipschitz Forward Backward Forward (M+LFBF) algorithm which was recently proposed in [15] offering both good performance and robustness to numerical errors. The convergence of this algorithm to a solution to (28) is guaranteed (under weak conditions) and its structure makes it easily implementable on highly parallel architectures.

6. SIMULATION RESULTS

Deriving performance bounds in compressed sensing in the presence of Poisson noise has been recently shown to be a challenging problem [16]. We subsequently provide some experiments in the Poisson context.

Figure 1 shows the reconstruction results for an electron microscopy image of size $N = 128 \times 128$ ($\mathcal{H} = \mathbb{R}^N$) degraded by a randomly decimated blur (uniform blur of size 3×3 and approximately 40% of missing data, that leads to $K = 9829$) and by Poisson noise with scaling parameter 0.5. We aim at evaluating the reconstruction results obtained when $C = [0, 255]^N$, and the data fidelity constraint related to the Poisson likelihood g is approximated as described in Section 5 with $M = 9$.

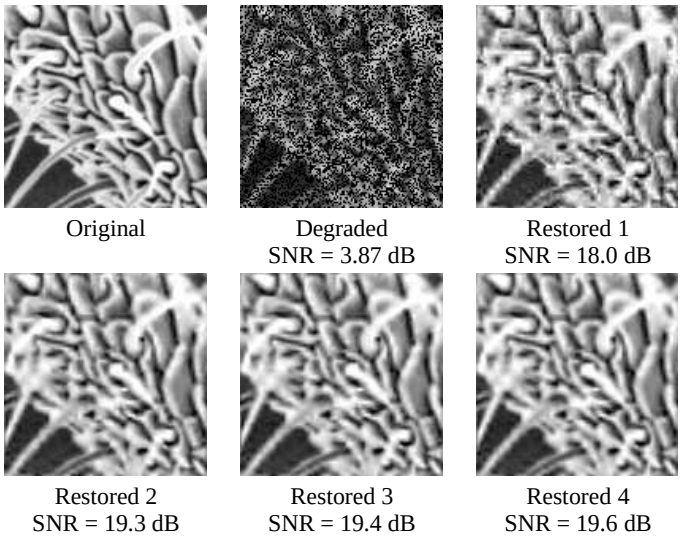

Fig. 1. Reconstruction results.

The *Restored 1* image displays the results we obtain when F is a wavelet transform using symmlet filters of length 6 applied over 2 resolution levels and the classical ℓ_1 criterion is employed. The images in the second row present the obtained results when F is a dual-tree transform (DTT) [17] using symmlet filters of length 6 applied over 2 resolution levels. Not only the use of a redundant frame allows us to reach higher SNR values, but the visual quality is significantly improved. *Restored 2* corresponds to the classical ℓ_1 cost function. *Restored 3* results from the minimization of a block sparsity measure for blocks gathering primal and dual DTT coefficients. *Restored 4* shows the good performance of the proposed approach when spatially overlapping blocks of size 2×2 are employed for each tree (primal or dual) separately

7. CONCLUSION

In this paper, we have proposed a convex optimization approach for solving cosparse modelling problems under flexible convex constraints. Our method is grounded on the use of recent proximal algorithms and a novel epigraphical projection technique. In the presented simulation results, this approach has been applied to a reconstruction problem involving data corrupted with Poisson noise. Other applications can be considered where the ability to take into account an arbitrary convex data fidelity constraint may be fruitful.

8. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projection : universal encoding strategies," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[3] S. Nam, M. Davies, M. Elad, and R. Gribonval, "Cosparse analysis modeling - uniqueness and algorithms," in *Int. Conf. on Acoust., Speech and Sig. Proc.*, Prague, Czech Republic, May 22 - 27, 2011, pp. 5804–5807.

[4] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. on Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[5] J.-J. Fuchs, "Extension of the global matched filter to structured groups of atoms: Application to harmonic signals.," in *Int. Conf. on Acoust., Speech and Sig. Proc.*, Prague, Czech Republic, May 22 - 27, 2011, pp. 3996–3999.

[6] G. Peyré and J. Fadili, "Group sparsity with overlapping partition functions," in *Proc. Eur. Sig. and Image Proc. Conference*, Barcelona, Spain, Aug. 29 – Sept. 2, 2011, pp. x+5.

[7] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *Technical report*, 2011, Preprint Hal-00621245.

[8] P. L. Combettes, "Inconsistent signal feasibility problems : least-squares solutions in a product space," *IEEE Trans. on Signal Proc.*, vol. 42, no. 11, pp. 2955–2966, Nov. 1994.

[9] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 97–123, Jan. 2011.

[10] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer-Verlag, New York, 2010.

[11] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. on Image Proc.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[12] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Math. Programm.*, vol. 64, pp. 81–101, 1994.

[13] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[14] L. M. Briceño-Arias and P. L. Combettes, "A monotone+skew splitting model for composite monotone inclusions in duality," *SIAM J. Optim.*, 2011, To appear, http://arxiv.org/abs/1011.5517.

[15] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, 2011, To appear, http://arxiv.org/abs/1107.0081.

[16] M. Raginsky, S. Jafarpour, Z. Harmany, R. Marcia, R. Willett, and R. Calderbank, "Performance bounds for expander-based compressed sensing in Poisson noise," *IEEE Trans. on Signal Proc.*, 2011, To appear, http://arxiv.org/abs/1007.2377.

[17] C. Chaux, L. Duval, and J.-C. Pesquet, "Image analysis using a dual-tree M-band wavelet transform," *IEEE Trans. on Image Proc.*, vol. 15, no. 8, pp. 2397–2412, Aug. 2006.