Optimization of a Geman-McClure like criterion for sparse signal deconvolution

Marc Castella, Jean-Christophe Pesquet

Optimization of a Geman-McClure Like Criterion for Sparse Signal Deconvolution

Marc Castella
SAMOVAR, Télécom SudParis,
CNRS, Université Paris-Saclay,
9 rue Charles Fourier - 91011 Evry Cedex, France
Email: marc.castella@telecom-sudparis.eu

Jean-Christophe Pesquet
Laboratoire d'Informatique Gaspard Monge, CNRS
Université Paris-Est
Champs sur Marne, France
Email: jean-christophe.pesquet@univ-paris-est.fr

***Abstract*—This paper deals with the problem of recovering a sparse unknown signal from a set of observations. The latter are obtained by convolution of the original signal and corruption with additive noise. We tackle the problem by minimizing a least-squares fit criterion penalized by a Geman-McClure like potential. The resulting criterion is a rational function, which makes it possible to formulate its minimization as a generalized problem of moments for which a hierarchy of semidefinite programming relaxations can be proposed. These convex relaxations yield a monotone sequence of values which converges to the global optimum. To overcome the computational limitations due to the large number of involved variables, a stochastic block-coordinate descent method is proposed. The algorithm has been implemented and shows promising results.**

I. Introduction

Many signal processing problems are undetermined in the sense that, from the available observations, it is not possible to infer unambiguously the signal of interest. The only way of circumventing this difficulty consists of incorporating prior information on the sought solution. In particular, the sparsity (possibly in some appropriate representation) is one of the most standard assumptions which can be made on the target signal in many situations of practical interest. A large literature has been devoted to signal/image recovery of sparse signals in connection with recent works on compressive sensing [1]. When the observations are obtained through a linear degradation model and some noise corruption process, many efforts have been undertaken in order to propose variational formulations of the problem. In such formulations, a sparsity measure is used in conjunction with a data fidelity term, such as a least squares criterion. The former can be introduced either as a penalization or under a constrained form, the two being related through Lagrangian duality under suitable conditions.

The natural sparsity measure is the ℓ_0 pseudo-norm which basically counts the number of nonzero components in the signal. It leads however to untractable NP-hard problems and to optimization difficulties due to the presence of many local minima [2]. Let us also mention the existence of iterative hard thresholding (IHT) algorithms, which can be quite effective in some cases, while having a low complexity [3]. These algorithms can be viewed as instances of the forward-backward (FB) iteration in the nonconvex case [4], [5]. Stochastic block-coordinate versions of IHT have also been recently proposed [6] and are related to existing works on block-coordinate FB algorithms [7], [8]. Nonetheless, for all these algorithms in general, convergence can be expected only to a local minimizer.

To improve the numerical performance, surrogates for the ℓ_0 cost function have been proposed. A well-known convex relaxation of this function is the ℓ_1 norm, yielding iterative soft thresholding methods [9], whose convergence is guaranteed. More generally, extensions of the ℓ_1 norm lead to so-called proximal thresholders which can be employed in provably convergent convex optimization schemes [4]. Although in some favorable scenarios the use of the ℓ_1 norm can be shown to be optimal for recovering a sparse signal [10], it is often suboptimal in terms of estimation of the support of its nonzero components and it introduces a bias in the estimation of their amplitudes. These drawbacks may be alleviated by making use of reweighted ℓ_1 minimization techniques [11]. Using an ℓ_1/ℓ_2 penalty may also lead to some improvements [12], [13].

Another kind of surrogates for the ℓ_0 pseudo-norm is provided by smoothed versions of the ℓ_1 or ℓ_0 function [14], [15]. In particular, one may be interested in sparsity measures of the form

$$(x_t)_{1\leq t\leq T} \in \mathbb{R}^T \mapsto \sum_{t=1}^{T} \psi_\delta(x_t) \ ,$$

where $\psi_\delta \colon \mathbb{R} \to \mathbb{R}$ is differentiable and $\delta \in]0,+\infty[$ is a smoothing parameter. Provided that $\psi_\delta(\sqrt{\cdot})$ is concave on $[0,+\infty[$, a quadratic tangent function can be derived, which makes efficient majorization-minimization (MM) strategies usable for optimizing penalized criteria built from this function (see [16] for more details). In addition if, for every $\xi \in \mathbb{R}$, $\lim_{\delta\to 0} \psi_\delta(\xi) = \chi_{\mathbb{R}\setminus\{0\}}(\xi)$ where $\chi_{\mathbb{R}\setminus\{0\}}(\xi) = 0$ when $\xi = 0$ and 1 otherwise, then the solution to the ℓ_0 penalized problem is recovered asymptotically as $\delta \to 0$ (under some technical assumptions) [14]. Among the class of possible smoothed ℓ_0 functions, the Geman-McClure $\ell_2 - \ell_0$ potential was observed to give good results in a number of applications [17], [14], [18]. It corresponds to the following choice for the function ψ_δ:

$$(\forall \xi \in \mathbb{R}) \qquad \psi_\delta(\xi) = \frac{\xi^2}{\delta^2 + \xi^2} \ . \tag{1}$$

Although efficient MM algorithms allow us to minimize penalized problems involving this function, they can get trapped by undesirable local minima due to the nonconvexity of the criterion. Note also that, when the signal to be recovered has positive values, a simplified form of (1) can be used:

$$(\forall \xi \in [0, +\infty[) \qquad \psi_\delta(\xi) = \frac{\xi}{\delta + \xi} \; . \tag{2}$$

Adding a penalization term such as (1) or (2) to a least squares criterion yields a rational objective function. Interestingly, we can take advantage of this fact through dedicated methods proposed in the optimization community [19], [20], [21], [22]. In these approaches, the minimization is recast as a problem of moments, for which a hierarchy of semidefinite positive (SDP) relaxations provides asymptotically an exact solution.

We investigate here the potential offered by these rational optimization methods for sparse signal deconvolution. Our method is based on recent developments in the field, providing theoretical guarantees of convergence to a global minimizer. In the present state of research, these methods are restricted to small or medium size problems and one of the main difficulties which we address is the large number of variables which have to be optimized. A new stochastic block-coordinate method will be proposed for this purpose.

The remainder of the paper is organized as follows. The convolutive model is introduced in Section II, as well as the associated variational formulation. Section III describes the optimization method used here, while Section IV presents the proposed stochastic block-coordinate strategy. A simulation example is shown in Section V and some concluding remarks are given in Section VI.

II. Model and criterion

A. Sparse signal model

We consider the problem of recovering a signal $(x_t)_{t\in\mathbb{Z}}$ which is assumed to be sparse: here, we simply assume that $x_t \neq 0$ only for a few indices t. Additionally, it is assumed that, for every $t \in \mathbb{R}$, $x_t \geq 0$. The signal $(x_t)_{t\in\mathbb{Z}}$ is unknown and the following real-valued observations $(y_t)_{1\leq t\leq T}$ are available:

$$(\forall t \in \{1,\ldots,T\}) \quad y_t = h_t \star x_t + n_t \; ,$$

where $\star$ denotes the convolution by the filter with impulse response $(h_t)_t$ and $\big(n_t\big)_{1\leq t\leq T}$ is an additive random independently and identically distributed (i.i.d.) noise. When the convolution filter has a finite impulse response (FIR) and cyclic boundary conditions are assumed, the above model can be rewritten as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \; ,$$

where $\mathbf{H}$ is a circulant Toeplitz matrix and $\mathbf{y}$, $\mathbf{x}$, $\mathbf{n}$ are $T \times 1$ column vectors containing the respective samples of the observations, unknown signal, and noise.

B. Criterion for recovery

As explained in the introduction, a classical approach for estimating $\mathbf{x} := (x_1,\ldots,x_T)^\top \in [0,+\infty[^T$ consists of minimizing a penalized criterion which in our case reads:

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda \sum_{t=1}^{T} \frac{x_t}{\delta + x_t} \; , \tag{3}$$

where λ and δ are positive parameters. The estimated signal is then $\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in[0,+\infty[^T} \mathcal{J}(\mathbf{x})$, where the minimization is performed over the feasible set $[0, +\infty[^T$. Note that the penalty term (2) has been chosen because of the nonnegativity assumption. However, the approach proposed in the paper remains valid when there is no such constraint and the penalization given by (1) is employed.

III. Minimization of a sum of rational functions

With an obvious notation, Criterion (3) can be expressed under the form:

$$\mathcal{J}(\mathbf{x}) = p_0(\mathbf{x}) + \sum_{t=1}^{T} \frac{p(x_t)}{q(x_t)} \tag{4}$$

and the problem to address then reads:

$$\mathcal{J}^\star := \inf_{\mathbf{x}\in\mathbf{K}} \mathcal{J}(\mathbf{x}) \; . \tag{5}$$

For technical reasons, we make the following assumption, which is easily satisfied when one knows an upper bound B on the variables $(x_t)_{1\leq t\leq T}$: the minimization set $\mathbf{K}$ is compact and decomposes as $\mathbf{K} = K_1 \times \cdots \times K_T$. The K_t's are here identical and are defined for every $t \in \{1,\ldots,T\}$ by polynomial inequalities which read for simplicity $K_t = \{x_t \in \mathbb{R} \mid g_t(x_t) \geq 0\}$ with $g_t(x_t) = x_t(B - x_t)$.

A. Generalized problem of moments

Let $\mathcal{M}(\mathbf{K})$ (resp. $\mathcal{M}(K_t)$) be the space of finite Borel measures supported on $\mathbf{K}$ (resp. K_t). In [23], the following infinite dimensional optimization problem is introduced:

$$\begin{aligned}
\mathcal{P}^\star := \inf & \int_{\mathbf{K}} p_0(\mathbf{x})\, d\mu_0(\mathbf{x}) + \sum_{t=1}^{T} \int_{K_t} p(x_t)\, d\mu_t(x_t) \\
\text{s.t.} & \int_{\mathbf{K}} d\mu_0(\mathbf{x}) = 1 \;\text{ and } (\forall \alpha \in \mathbb{N})(\forall t \in \{1\ldots,T\}) : \\
& \int_{K_t} x_t^\alpha q(x_t)\, d\mu_t(x_t) = \int_{\mathbf{K}} x_t^\alpha \, d\mu_0(\mathbf{x}) \; ,
\end{aligned}$$

where the variables are measures $\mu_t, t \in \{0,\ldots,T\}$, with $\mu_0 \in \mathcal{M}(\mathbf{K})$ and $\mu_t \in \mathcal{M}(K_t)$ for $t \geq 1$. It can been shown that $\mathcal{P}^\star = \mathcal{J}^\star$ under the assumption that $\mathbf{K}$ is compact. This can be accounted for by the fact that any global optimum point $\mathbf{x}^\star$ of (5) corresponds to the following set of Dirac measures $\mu_0 = \delta_{\mathbf{x}^\star}$ and $\mu_t = q(x_t^\star)^{-1}\delta_{x_t^\star}$, for every $t \in \{1,\ldots,T\}$. This fact was first presented in the context of polynomial optimization in [19]. It must however be emphasized that the rational criterion (4) has very high numerator and denominator degrees when reducing it to a single fraction, which does not allow us to use the methods in [20] or [22] for optimizing rational functions, and which was the motivation for the work in [23].

B. A hierarchy of SDP relaxation

For numerical tractability, the infinite dimensional optimization problem $\mathcal{P}^\star$ needs to be relaxed to a finite dimensional SDP: the first ingredient is to represent the different measures $\mu_0, \mu_1, \ldots, \mu_T$ of problem $\mathcal{P}^\star$ by their respective moment sequences $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_T$. Since μ_0 is a measure on $\mathbf{K} \subset [0, +\infty[^T$, $\mathbf{y}_0$ is indexed with multi-indices in $\mathbb{N}^T$ corresponding to the monomial exponents in the canonical basis $(\mathbf{x}^{\boldsymbol{\alpha}})$ of $\mathbb{R}[\mathbf{x}]$. Conversely, for every $t \geq 1$, the measure μ_t is defined on $K_t \subset \mathbb{R}$, and $\mathbf{y}_t$ is indexed by a number in $\mathbb{N}$. For any moment sequence $\mathbf{y}$, we define the following linear functional, which replaces any monomial in the polynomial $f \in \mathbb{R}[\mathbf{x}]$ by the corresponding moment value in $\mathbf{y}$:

$$\begin{aligned} L_{\mathbf{y}} :\ & \mathbb{R}[\mathbf{x}] \to \mathbb{R} \\ & f = \sum f_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} \mapsto \sum f_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}} \ . \end{aligned}$$

For any order $k \in \mathbb{N}$ and for multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta}$ of order $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_n \leq k$ and $|\boldsymbol{\beta}| \leq k$, the *moment matrix* of $\mathbf{y}$ is defined by

$$[M_k(\mathbf{y})]_{\boldsymbol{\alpha},\boldsymbol{\beta}} := y_{\boldsymbol{\alpha}+\boldsymbol{\beta}} \ ,$$

and for a given polynomial $g \in \mathbb{R}[\mathbf{x}]$, the *localizing matrix* associated to g and $\mathbf{y}$ is

$$[M_k(g\mathbf{y})]_{\boldsymbol{\alpha},\boldsymbol{\beta}} := \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}} y_{\boldsymbol{\gamma}+\boldsymbol{\alpha}+\boldsymbol{\beta}} \ .$$

Finally, define $r_t := \lceil (\deg g_t)/2 \rceil$. In [23], a hierarchy of sparse SDP relaxations has been proposed. Defining $k \in \mathbb{N}$ as the order of the relaxation in the hierarchy, the latter reads:

$$\begin{aligned} \mathcal{P}_k^\star := \inf\ & L_{\mathbf{y}_0}(p_0) + \sum_{i=1}^{T} L_{\mathbf{y}_i}(p) \\ \text{s.t. }\ & M_k(\mathbf{y}_0) \succeq 0, L_{\mathbf{y}_0}(1) = 1 \text{ and } (\forall t \in \{1, \dots, T\}) : \\ & M_k(\mathbf{y}_t) \succeq 0 \\ & M_{k-r_t}(g_t \mathbf{y}_0) \succeq 0 \\ & M_{k-r_t}(g_t \mathbf{y}_t) \succeq 0 \\ & L_{\mathbf{y}_t}(x_t^\alpha q(x_t)) = L_{\mathbf{y}_0}(x_t^\alpha) \text{ for } \alpha + \deg q \leq 2k \ . \end{aligned}$$

It has been proved in [23] that the associated monotone sequence of optimal values for the above hierarchy of SDP relaxations converges to the global optimum, that is

$$\mathcal{P}_k^\star \uparrow \mathcal{J}^* \text{ as } k \to \infty \ .$$

In addition, under certain rank conditions, global minimizers of (5) can be extracted [24]. Fortunately, low relaxation orders often provide satisfactory results.

IV. Block-coordinate optimization

Although very appealing, the approach described in Section III has a major drawback: the numbers of variables in the moment sequences $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_T$ is large for even small values of the relaxation order k and of the number of samples T (e.g. SDP of size 5650 for $T = 100$ and $k = 1$). As a consequence, we propose to perform the optimization with respect to $\mathbf{x}$ using a stochastic block-coordinate descent method. Let $\tilde{\mathbf{x}}$ be an $N \times 1$ column vector containing a subset of N components of $\mathbf{x}$ ($N \leq T$) and let $\bar{\mathbf{x}}$ be the $(T-N) \times 1$ column vector containing the remaining components of $\mathbf{x}$. We partition similarly the columns of $\mathbf{H}$, and define $\tilde{\mathbf{H}}$ (resp. $\bar{\mathbf{H}}$) the $N \times T$ (resp. $(T-N) \times T$) matrices obtained from $\mathbf{H}$ such that $\mathbf{H}\mathbf{x} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \bar{\mathbf{H}}\bar{\mathbf{x}}$. We then have $\mathcal{J}(\mathbf{x}) = \tilde{\mathcal{J}}(\tilde{\mathbf{x}}) + \text{const.}$ where

$$\tilde{\mathcal{J}}(\tilde{\mathbf{x}}) := \|\mathbf{y} - \tilde{\mathbf{H}}\tilde{\mathbf{x}} - \bar{\mathbf{H}}\bar{\mathbf{x}}\|^2 + \lambda \sum_{t=1}^{N} \frac{\tilde{x}_t}{\delta + \tilde{x}_t} \ ,$$

and the constant depends on $\bar{\mathbf{x}}$ only. With an obvious notation, the above criterion takes the form (4) and the method from Section III-B can then be employed to optimize $\tilde{\mathcal{J}}(\tilde{\mathbf{x}})$. After initializing $\mathbf{x}$ to a value $\mathbf{x}_{\text{ini}}$, our optimization procedure is thus the iteration of the following steps:

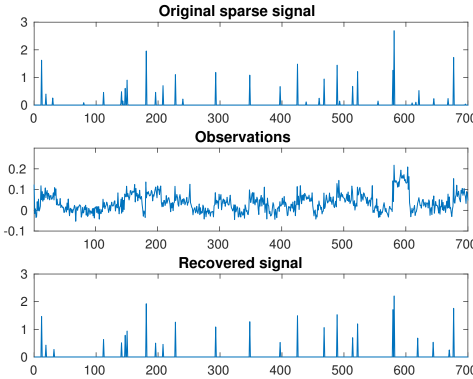

Fig. 1. Typical observation and unknown sparse signal

- draw randomly N indices $t_1, \dots, t_N$ in $\{1, \dots, T\}$ and define $\tilde{\mathbf{x}} := (x_{t_1}, \dots, x_{t_N})^\top$.
- optimize $\tilde{\mathcal{J}}(\tilde{\mathbf{x}})$, that is
  - build the corresponding relaxation $\mathcal{P}_k^\star$
  - solve the SDP problem and extract its solution.
- update $(x_{t_1}, \dots, x_{t_N})^\top \leftarrow \arg\min_{\tilde{\mathbf{x}}} \tilde{\mathcal{J}}(\tilde{\mathbf{x}})$.

V. Simulations

A. Software and implementation

The hierarchy $\mathcal{P}_k^\star$ of SDP relaxations of the generalized problem of moments $\mathcal{P}^\star$ can be easily built and solved. Indeed, the Matlab software package GloptiPoly3 [25] allows one to build the hierarchy in a user friendly way. GloptiPoly3 can then solve it by calling one of the publicly available SDP solvers. In our simulations, we used the solver SeDuMi [26]. Finally, GloptiPoly3 can return the solution found for the SDP relaxation.

B. Illustration results

We have generated $T = 700$ samples of a sparse signal according to an i.i.d. process. Each component x_t of $\mathbf{x}$ has a probability distribution given by $0.95\delta_0 + 0.05\mathcal{N}^+(0,1)$ where $\mathcal{N}^+(0,1)$ is a positive truncated Gaussian law. The convolution filter is a low-pass FIR with length 25. The noise standard-deviation is set to $\sigma = 0.02$. A typical plot of the unknown $\mathbf{x}$ and the observed $\mathbf{y}$ is given in Figure 1. The corresponding recovered signal is shown at bottom. Such a scenario is likely to occur in several applications such as seismic deconvolution or spectroscopy. For comparison, we tried to perform the deconvolution on the same set of samples by minimizing the ℓ_1 penalized criterion $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda_{\ell_1} \sum_{t=1}^{T} |x_t|$ subject to the positivity constraint $\mathbf{x} \in [0, +\infty[^T$, where $\lambda_{\ell_1} \in]0, +\infty[$. We also implemented an IHT method. The parameter values have been set empirically to get the best possible results ($\lambda = 5 \times 10^{-3}$, $\delta = 2 \times 10^{-1}$). Finally, to confirm that a global minimum is reached, different initializations have been tried: zero ($\mathbf{x}_{\text{ini}} = \mathbf{0}$), the observation vector ($\mathbf{x}_{\text{ini}} = \mathbf{y}$), the result of ℓ_1 penalization ($\mathbf{x}_{\text{ini}} = \mathbf{x}_{\ell_1}$), values randomly drawn in $[0,1]$ ($\mathbf{x}_{\text{ini}} = \mathbf{x}_{\text{rand}}$) and the true value ($\mathbf{x}_{\text{ini}} = \mathbf{x}_{\text{true}}$). The last initialization is of no use in practice, but provides an interesting reference.

On Figure 2 we plot the objective value $\mathcal{J}(\mathbf{x})$ (top) and the mean square error (MSE) (bottom) with respect to the

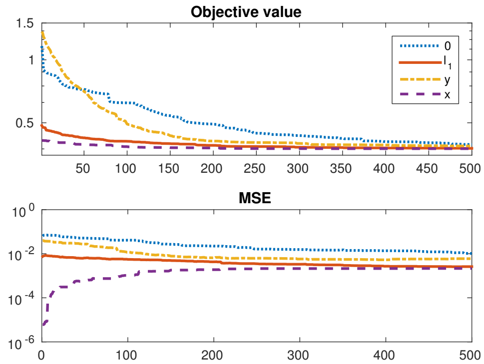

Fig. 2. Criterion $\mathcal{J}(\mathbf{x})$ and MSE versus iterations. See the legend box for the different initialization values of $\mathbf{x}_{\text{ini}}$.

TABLE I. FINAL VALUES OF THE OBJECTIVE $\mathcal{J}(\mathbf{x})$ AND MSE FOR OUR METHOD, IHT AND THE ℓ_1 PENALIZATION.

	Proposed method		IHT		ℓ_1 penalization	
$\mathbf{x}_{\text{ini}}$	$\mathcal{J}(\mathbf{x})$	MSE	$\mathcal{J}(\mathbf{x})$	MSE	$\mathcal{J}(\mathbf{x})$	MSE
$\mathbf{0}$	0.3758	0.0035	1.1951	0.0434		
$\mathbf{x}_{\ell_1}$	0.3760	0.0022	0.4030	0.0048		
$\mathbf{y}$	0.3777	0.0032	0.4452	0.0063	0.5394	0.0068
$\mathbf{x}_{\text{true}}$	0.3760	0.0022	0.3877	0.0012		
$\mathbf{x}_{\text{rand}}$	0.3758	0.0035	0.5693	0.0298		

sought signal versus the block-coordinate iteration number. We performed 5000 iterations but only the 500 first ones have been plotted. We clearly observe that the objective decreases and converges to the same minimal value for *any* initialization point $\mathbf{x}_{\text{ini}}$: this advocates in favor of a global convergence. The final values of $\mathcal{J}(\mathbf{x})$ are given in Table I for our method, IHT and ℓ_1 penalization. Clearly, our method finds close optimal values for any initialization, whereas IHT is sensitive to local minima. Similarly, the results concerning the MSE appear to be quite consistent.

VI. Conclusion

The deconvolution of a sparse signal has been considered through the minimization of a least squares criterion penalized by a Geman-McClure like potential. The resulting objective is non convex but rational. For such minimization, we have employed recent methodological tools offering theoretical guarantee of global convergence. Due to the important number of variables, we have proposed to split the problem into a sequence of blockwise optimization steps. Very promising experimental results have been obtained.

References

[1] E. J. Candès and M. B. M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[2] M. Nikolova, "Description of the minimizers of least squares regularized with ℓ_0 norm. uniqueness of the global minimizer," *SIAM J. Imaging Sci.*, vol. 6, no. 2, pp. 904–937, 2013.

[3] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, 2008.

[4] 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. New York: Springer-Verlag, 2010, pp. 185–212.

[5] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized Gauss-Seidel methods," *Math. Programm.*, vol. 137, no. 1-2, pp. 1–39, Feb. 2013.

[6] A. Patrascu and I. Necoara, "Random coordinate descent methods for ℓ_0 regularized convex optimization," *IEEE Trans. Automat. Contr.*, vol. 60, no. 7, pp. 1811–1824, Jul. 2015.

[7] P. L. Combettes and J.-C. Pesquet, "Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping," *SIAM J. Optim.*, vol. 25, no. 2, pp. 1221–1248, July 2015.

[8] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," Tech. Rep., 2015. [Online]. Available: http://www.optimization-online.org/DB_HTML/2013/12/4178.html

[9] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[10] M. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.

[11] E. J. Candés, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[12] N. Hurley and G. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[13] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet, "Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization," *Signal Process. Lett.*, vol. 22, no. 5, pp. 539–543, May 2015.

[14] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 563–591, 2013.

[15] E. Soubies, L. Blanc-Fraud, and G. Aubert, "A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem," *SIAM J. Imaging Sci.*, vol. 8, no. 3, pp. 1607–1639, 2015.

[16] E. Chouzenoux, J. Idier, and S. Moussaoui, "A majorize-minimize subspace strategy for subspace optimization applied to image restoration," *IEEE Trans. Image Process.*, vol. 20, no. 18, pp. 1517–1528, Jun. 2011.

[17] S. Geman and D. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Statist. Comput. Section Amer. Statist. Association*, 1985, pp. 12–18.

[18] A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu, and S. Ciochina, "A majorize-minimize memory gradient method for complex-valued inverse problem," *Signal Process.*, vol. 103, pp. 285–295, Oct. 2014, special issue on Image Restoration and Enhancement: Recent Advances and Applications.

[19] J.-B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optim.*, vol. 11, no. 3, pp. 796–817, 2001.

[20] D. Jibetean and E. de Klerk, "Global optimization of rational functions: a semidefinite approach," *Math. Progr. (Ser. A)*, no. 106, 2005.

[21] M. Laurent, "Sum of squares, moment matrices and optimization over polynomials," in *Emerging Applications of Algebraic Geometry*, ser. IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant, Eds. Springer, 2009, vol. 149, pp. 157–270.

[22] J.-B. Lasserre, *Moments, Positive Polynomials and Their Applications*, ser. Optimization Series. Imperial College Press, 2010, vol. 1.

[23] F. Bugarin, D. Henrion, and J.-B. Lasserre, "Minimizing the sum of many rational functions," LAAS-CNRS Research Report 11100, Tech. Rep., February 2011, updated in Sep. 2013 and Dec. 2014, submitted to Mathematical Programming Computation.

[24] D. Henrion and J.-B. Lasserre, "Detecting global optimality and extracting solutions in GloptiPoly," Tech. Rep., 2005.

[25] D. Henrion, J.-B. Lasserre, and J. Löfberg, "Gloptipoly3: moments, optimization and semidefinite programming," *Optimization methods and software*, vol. 24, no. 4-5, pp. 761–779, 2009.

[26] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.