Theory and Applications of Mathematics & Computer Science

(ISSN 2067-2764, EISSN 2247-6202)
http://www.uav.ro/applications/se/journal/index.php/tamcs

Theory and Applications of Mathematics & Computer Science 5 (1) (2015) 1–19

Sparse and Robust Signal Reconstruction

Sandra V. B. Jardim*

Tomar Polytechnic Institute, Department of Information Technology, Quinta do Contador, Estrada da Serra - Tomar, Portugal.

Abstract

Many problems in signal processing and statistical inference are based on finding a sparse solution to an undetermined linear system. The reference approach to this problem of finding sparse signal representations, on overcomplete dictionaries, leads to convex unconstrained optimization problems, with a quadratic term ℓ_2, for the adjustment to the observed signal, and a coefficient vector ℓ_1-norm. This work focus the development and experimental analysis of an algorithm for the solution of ℓ_q-ℓ_p optimization problems, where $p \in]0, 1] \wedge q \in [1, 2]$, of which ℓ_2-ℓ_1 is an instance. The developed algorithm belongs to the majorization-minimization class, where the solution of the problem is given by the minimization of a progression of majorizers of the original function. Each iteration corresponds to the solution of an ℓ_2-ℓ_1 problem, solved by the projected gradient algorithm. When tested on synthetic data and image reconstruction problems, the results shows a good performance of the implemented algorithm, both in compressed sensing and signal restoration scenarios.

Keywords: Sparse signal representation, Convex relaxation, ℓ_2-ℓ_1 optimization, Compressed sensing, Majorization-minimization algorithms, Quadratic programming, Gradient projection algorithms.
2010 MSC: 93C42, 94A12.

1. Introduction

In general, sparse approximation problems have been of great interest given its wide applicability both in signal and image processing field as in statistical inference contexts, where many of the problems to be solved involve the undetermined linear systems sparse solutions determination.

The literature on sparsity optimization is rapidly increasing (see (Zarzer, 2009; Donoho, 2006; Candes & Tao, 2005; Wright *et al.*, 2009) and references therein). More recently sparsity techniques are also receiving increased attention in the optimal control community (Stadler, 2009; Casas *et al.*, 2012; Herzog *et al.*, 2012).

Given an input signal $\mathbf{y}$, sparse approximation problems resolution aims an approximated signal determination $\mathbf{x}$ through a linear combination of elementary signals, which are, for several

*Corresponding author
Email address: sandra.jardim@ipt.pt (Sandra V. B. Jardim)

current applications, extracted from a set of signals not necessarily lineary independent. A preference for sparse linear combinations is imposed by penalizing nonzero coefficients. The most common penalty is the number of elementary signals that participate in the approximation. According to the context in which they operate and the objective to be achieved, sparse approximation problems can be formulated in different ways. In this sense, the problem domain must justify the linear model, the elementary signals choice and the sparsity criterion to be adopted.

On account of the combinatorial nature of the sparse approximation problems, which is due to the presence of the *quasi*-norm ℓ_0 of the coefficients vector to be estimated, these problems have a difficult computational resolution. In general, these optimization problems are NP-hard problems (Davis *et al.*, 1997; Natarajan, 1995). One of the most common approach to overcome this difficulty is the convex relaxation, introduced by Claerbourt et al. (Claerbout & Muir, 1973), of the original problem, where the *quasi*-norm ℓ_0 is replaced by the norm ℓ_1, which is a convex function. A classic example of this kind of problems is the determination of sparse representations on overcomplete dictionaries, where the reference approach leads to unconstraint convex optimization problems, which involves a quadratic term ℓ_2 of adjustment to the observed signal **y**, and a ℓ_1 norm of the coefficient vector to be estimated **x**. In this sense, it is notorious the interest shown by the scientific community in the development of methods leading to the resolution of the unconstraint convex optimization problem

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \phi\,\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1, \tag{1.1}$$

where ϕ represents the overcomplete dictionary synthesis matrix; if $\mathbf{x} \in \Re^n$ and $\mathbf{y} \in \Re^k$, ϕ is a $k \times n$ matrix. The nonnegative parameter λ states a compromise between the approximation mean squared error and his sparsity level.

The above optimization problem, and related ones, arises in several applications, such as the *Basis Pursuit* and *Basis Pursuit Denoising* criterions (Chen *et al.*, 2001) and *Compressed Sensing* (Donoho, 2006).

The optimization problem represented in (1.1) is an instance of the general class of the optimization problems ℓ_q - ℓ_p, where q and p can assume values in the range]0, 2]. It is important to stress that the data term generalization to a ℓ_q norm, instead of the ℓ_2 norm, gives to the approximation criterion statistical strength features (when $q < 2$) (Huber & Ronchetti, 2009) , making it less permeable to spurious observations (*outliers*). On the other hand, when it comes to generalization of the coefficient term to be estimated to a ℓ_p norm, instead a ℓ_1 norm, and considering $p < 1$, we walk toward the original combinatorial problem resolution.

In this paper is presented an algorithm that aims the resolution of the general class optimization problems ℓ_q -ℓ_p, where $p \in]0, 1] \wedge q \in [1, 2]$. The developed algorithm belong to the majorization-minimization class (Hunter & Lange, 2004), where the problem is solved in an iterative way, through the minimization of a majorizers sequence of the original function. Each upper bound function corresponds to a ℓ_2 - ℓ_1 problem, where each one of these problems is formulated as a quadratic programming problem and solved through the gradient projection algorithm (Figueiredo *et al.*, 2007*b*).

2. Generalized Optimization Problem

The unconstrained optimization problem represented in (1.1) is the convex relaxation of the subset selection problem, checking to be a major problem in many application areas. Due to its high applicability, there has been considerable effort made by the scientific community, regarding techniques and algorithms development for its resolution. Among the different proposed algorithms, are homotopy algorithms (Turlach, 2005; Efron *et al.*, 2004; Malioutov *et al.*, 2005), the ones based on interior-point methods (Turlach *et al.*, 2005; Chen *et al.*, 2001), the majorization-minimization class algorithms (Figueiredo & Nowak, 2003, 2005; Figueiredo *et al.*, 2007*a*) and the gradient projection algorithm (Figueiredo *et al.*, 2007*b*).

In this problem, the objective function composed by two terms, one of which being a quadratic term ℓ_2 of adjustment to the observed signal $\mathbf{y}$, and the other the ℓ_1 norm of the coefficient vector to estimate $\mathbf{x}$. Recall that the ℓ_1 norm arises by replacing the *quasi*-norm ℓ_0, at the convex relaxation of the original convex optimization problem.

As stated above, the optimization problem ℓ_2 - ℓ_1 is an instance of the optimization problems ℓ_q - ℓ_p general class, where $p \in]0,\ 1] \wedge q \in [1,\ 2]$.

Since the purpose of this work is to achieve the solution of the general class optimization problems ℓ_q - ℓ_p, for $p \in]0,\ 1] \wedge q \in [1,\ 2]$, let's consider the generalization of the unconstrained convex optimization problem (1.1), and define the function $L(\mathbf{x})$

$$L(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \phi\,\mathbf{x}\|_q^q + \lambda\|\mathbf{x}\|_p^p, \tag{2.1}$$

where $\mathbf{x} \in \Re^n$, $\mathbf{y} \in \Re^k$, ϕ is the synthesis matrix of a dictionary D (of dimension $k \times n$), $\lambda \geq 0$, $p \in]0,\ 1]$ and $q \in [1,\ 2]$.

3. Majorization-Minimization Method for the Resolution of the Generalized Optimization Problem

3.1. Objective Function

There are several algorithms for the resolution of the optimization problem (1.1), such as Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), Homotopy, Least Absolute Shrinkage and Selection Operator (LASSO) and Gradient Projection (GPSR). In this work is developed a majorization-minimization (MM) method, for the resolution of the optimization problem $\min_{\mathbf{x}} L(\mathbf{x})$, where the optimization problem (1.1) is solved using the Barzilai-Borwein Gradient Projection algorithm for sparse reconstrution (GPSR-BB) (Figueiredo *et al.*, 2007*b*). This choice results from an analysis of the results obtained for different algorithms, which can be found in (Jardim, 2008). Since GPSR-BB algorithm aims to solve the optimization problem (1.1), it is necessary to establish a relation between this and the optimization problem that results from the minimization of the function $L(\mathbf{x})$ (2.1).

We can observe that the function $L(\mathbf{x})$ is the sum of the functions $L_1(\mathbf{x})$ and $L_2(\mathbf{x})$, where

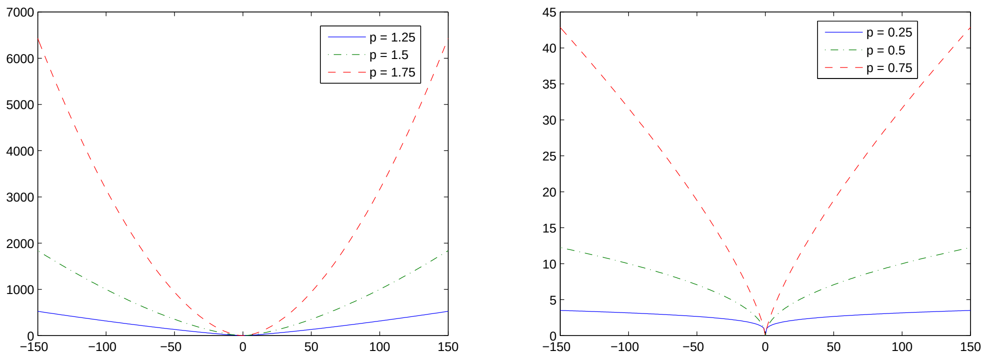
$$L_1(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \phi\,\mathbf{x}\|_q^q \quad \text{and} \tag{3.1}$$

$$L_2(\mathbf{x}) = \lambda \|\mathbf{x}\|_p^p. \tag{3.2}$$

Knowing that $\|w\|_r^r = \sum_i |w_i|^r$, we can define the function

$$f(z,\ p) = |z|^p. \tag{3.3}$$

Figures (1(a)) and (1(b)) are graphical representations of the function $f(z,\ p)$ for different values of p.



(a) $p = 1.75$, $p = 1.5$ and $p = 1.25$.

(b) $p = 0.75$, $p = 0.5$ and $p = 0.25$.

Figure 1. Function $f(z,\ p) = |z|^p$ for values of p greater and smaller than 1.

With this function it is possible to write (3.1) and (3.2) as

$$L_1(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{k} f\left(y_i - (\phi\mathbf{x})_i,\ q\right). \tag{3.4}$$

and

$$L_2(\mathbf{x}) = \lambda \sum_{i=1}^{n} f\left(x_i,\ p\right). \tag{3.5}$$

3.1.1. The Majorizer Function

By the analysis of the figures (1(a)) and (1(b)) we can verify that it is possible to determine for the function $f(\mathbf{z},\ p)$ (so to the $L_1(\mathbf{x})$ and $L_2(\mathbf{x})$ functions), and for some value $\mathbf{z}'$, majorizers functions. This opens the door to the use of majorization-minimization algorithms to the resolution of the optimization problem $\min_{\mathbf{x}} L(\mathbf{x})$, where $L(\mathbf{x})$ is the function given by (2.1). So, it is necessary to define a Q function such that

$$L(\mathbf{x}) \leq Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)}),\ \forall_{\mathbf{x}\neq\hat{\mathbf{x}}^{(t)}} \tag{3.6}$$

$$L(\hat{\mathbf{x}}^{(t)}) = Q(\hat{\mathbf{x}}^{(t)}|\hat{\mathbf{x}}^{(t)}), \tag{3.7}$$

i.e., $Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)})$ is a function of $\mathbf{x}$ that majorizes (*i.e.*, upper bounds) $L(\mathbf{x})$.

Recalling that $L(\mathbf{x}) = L_1(\mathbf{x}) + L_2(\mathbf{x})$, where $L_1(\mathbf{x})$ and $L_2(\mathbf{x})$ are given by (3.4) and (3.5) respectively, let's consider the majorizer functions

$$Q_1(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) \geq L_1(\mathbf{x}) \ \text{ and } \ Q_1(\hat{\mathbf{x}}^{(t)}|\hat{\mathbf{x}}^{(t)}) = L_1(\hat{\mathbf{x}}) \tag{3.8}$$

and

$$Q_2(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) \geq L_2(\mathbf{x}) \ \text{ and } \ Q_2(\hat{\mathbf{x}}^{(t)}|\hat{\mathbf{x}}^{(t)}) = L_2(\hat{\mathbf{x}}^{(t)}). \tag{3.9}$$

Given the MM algorithm properties (Hunter & Lange, 2004), we can define a function

$$Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) = Q_1(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) + Q_2(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) \ \text{ such that} \tag{3.10}$$

$$L(\mathbf{x}) \leq Q_1(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) + Q_2(\mathbf{x}|\hat{\mathbf{x}}^{(t)}), \tag{3.11}$$

verifying

$$L(\hat{\mathbf{x}}^{(t)}) = Q_1(\hat{\mathbf{x}}^{(t)}|\hat{\mathbf{x}}^{(t)}) + Q_2(\hat{\mathbf{x}}^{(t)}|\hat{\mathbf{x}}^{(t)}). \tag{3.12}$$

Due to (3.4), and assuming q a fixed value in the interval $[1, 2]$, $L_1(\mathbf{x})$ is an even and growing function with $|\mathbf{y} - \phi\,\mathbf{x}|$ (see figure (1(a))), having a slower growth, or equal case $q = 2$, than a quadratic function of $|(\mathbf{y} - \phi\,\mathbf{x})_i|$. So, it makes sense to use as majorizer function of $L_1(\mathbf{x})$ a quadratic function, which is also an even function, so without a linear term. In other words, we can use as $L_1(\mathbf{x})$ majorizer a function of the form $\frac{1}{2}\sum_i a_i(\mathbf{y} - \phi\,\mathbf{x})_i^2 + b_i$. The upper bound function can be used in a MM algorithm only if it verifies the key property, *i. e.*, the upper bound function must touch the function at the previous estimative. So it is necessary to determine a_i and b_i in order to

$$a_i(\mathbf{y} - \phi\,\mathbf{x})_i^2 + b_i \ \geq \ |(\mathbf{y} - \phi\,\mathbf{x})_i|^q, \forall_{\mathbf{x}} \tag{3.13}$$

$$a_i(\mathbf{y} - \phi\,\mathbf{x}')_i^2 + b_i \ = \ |(\mathbf{y} - \phi\,\mathbf{x}')_i|^q. \tag{3.14}$$

Let's consider the function $f(z, q)$ given by (3.3), and a quadratic majorizer such that

$$g_1(z, z') = a\,z^2 + b \text{ such that:}$$
$$g_1(z, z') \geq f(z, q)\ \forall_z \ \text{ and } \ g_1(z', z') = f(z', q). \tag{3.15}$$

Given that $f(z, q) = |z|^q$ we have, for $q \in [1, 2]$ and $z \neq 0$,

$$\frac{d\,f(z, q)}{dz} = q|z|^{(q-1)}\text{sign}(z). \tag{3.16}$$

In order to $g_1(z, z')$ be tangent to $f(z, q)$ at $z = z'$, we get

$$a = \frac{q}{2}|z'|^{(q-2)}. \tag{3.17}$$

Since we also want $f(z', q) = g_1(z', z')$, we have

$$b = \frac{2-q}{2}|z'|^{q}. \tag{3.18}$$

Finally, we can write the majorizer function

$$g_1(z,\ z') = \frac{q}{2}(z')^{(q-2)}z^2 + \frac{2-q}{2}|z'|^{q}. \tag{3.19}$$

Based on majorizer (3.19) we can write

$$\begin{aligned}\frac{1}{2}\|\mathbf{y}-\phi\,\mathbf{x}\|_q^q &= \frac{1}{2}\sum_i |(\mathbf{y}-\phi\,\mathbf{x})_i|^q \\ &\leq \frac{q}{4}\sum_i |(\mathbf{y}-\phi\,\mathbf{x}')_i|^{(q-2)}(\mathbf{y}-\phi\,\mathbf{x})_i^{\,2} \\ &+ \frac{2-q}{2}|(\mathbf{y}-\phi\,\mathbf{x}')_i|^q.\end{aligned} \tag{3.20}$$

Defining

$$\alpha_i^{(t)} = \frac{q}{2}\left(y_i - \sum_j \left(\phi_{ij}\,\hat{x}_j^{(t)}\right)\right)^{q-2}, \tag{3.21}$$

and

$$\beta_i^{(t)} = \frac{2-q}{2}\left|y_i - \sum_j \left(\phi_{ij}\,\hat{x}_j^{(t)}\right)\right|^{q}, \tag{3.22}$$

we have the majorizer $Q_1(\mathbf{x}|\hat{\mathbf{x}}^{(t)})$ given by

$$Q_1(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) = \frac{1}{2}\sum_i \left[\alpha_i^{(t)}\left(y_i - \sum_j \phi_{ij}\,x_j\right)^2 + \beta_i^{(t)}\right]. \tag{3.23}$$

Since $L(\mathbf{x})$ is a function consisting of two separable terms, for which can be defined different majorizer functions, the majorizer function for the term $L_2(\mathbf{x}) = \lambda\|\mathbf{x}\|_p^p$ does not necessarily have to be a quadratic one. In fact, what is desirable is that the majorizer to be adopt for the $L_2(\mathbf{x})$ function, when added to the majorizer defined for $L_1(\mathbf{x})$, leads to a function $Q(\mathbf{x}, \hat{\mathbf{x}}^{(t)})$ with a minimizer easy to find. So, a ℓ_1 majorizer is the natural choice for penalties ℓ_p, with $0 < p \leq 1$, since it is more tighter than a quadratic majorizer. Recalling that, for the upper bound function can be used in a MM algorithm it must touch the function at the previous estimate, it is necessary to determine the parameters c and d so that

$$|x|^p \leq c\,|x| + d,\ \forall_x \ \text{ and } \ |x'|^p \leq c\,|x'| + d.$$

Let's consider the function $f(x,\ p)$ given by (3.3), and a majorizer ℓ_1:

$$\begin{aligned} g_2(x,\ x') = c\,|x| + d \ \text{ such that:} \\ g_2(x,\ x') \geq f(x,\ p),\ \forall_x \ \text{ and } \ g_2(x',\ x') = f(x',\ p). \end{aligned} \tag{3.24}$$

Given that, for $p \in\ ]0,\ 1]$, and $x' \neq 0$

$$\frac{d f(x,\ p)}{dx}|_{x=x'} = \text{sign}(x')p|x'|^{(p-1)}, \tag{3.25}$$

we have

$$c = p|x'|^{(p-1)}, \tag{3.26}$$

and

$$d = (1-p)|x'|^p. \tag{3.27}$$

We can then write

$$g_2(x,\ x') = p|x'|^{(p-1)}|x| + (1-p)|x'|^p. \tag{3.28}$$

Defining

$$\delta_i^{(t)} = \begin{cases} p\,|\hat{x}_i^{(t)}|^{(p-1)}, & \text{if } \hat{x}_i^{(t)} \neq 0 \\ 1/eps\,^{a}, & \text{if } \hat{x}_i^{(t)} = 0 \end{cases} \tag{3.29}$$

and

$$\epsilon_i^{(t)} = (1-p)\,|\hat{x}_i^{(t)}|^p. \tag{3.30}$$

we have that the function $Q_2(\mathbf{x}|\hat{\mathbf{x}}^{(t)})$ can be written as

$$Q_2(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) = \lambda \sum_i \left[\delta_i^{(t)}\,|x_i| + \epsilon_i^{(t)}\right]. \tag{3.31}$$

Given (3.10) we have that

$$\begin{aligned} Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) \ &= \ \frac{1}{2}\sum_{i=1}^{k}\left[\alpha_i^{(t)}\left(y_i - \sum_j \phi_{ij}\,x_j\right)^2 + \beta_i^{(t)}\right] \\ &+ \ \lambda \sum_{i=1}^{n}\left[\delta_i^{(t)}\,|x_i| + \epsilon_i^{(t)}\right]. \end{aligned} \tag{3.32}$$

[a] eps is the distance from 1.0 to the next larger double precision number, that is eps with no arguments returns $2^{(-52)}$.

Since at each step of the MM algorithm, we perform a minimization in $\mathbf{x}$, the terms $\beta_i^{(t)}$ and $\epsilon_i^{(t)}$ can be ignored, so we have

$$\begin{aligned} Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) &= \frac{1}{2}\sum_i \left[\alpha_i^{(t)}\left(y_i - \sum_j \phi_{ij}\, x_j\right)^2\right] \\ &+ \lambda \sum_i \left[\delta_i^{(t)}\, |x_i|\right], \end{aligned} \tag{3.33}$$

or, in vectorial notation

$$Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) = \frac{1}{2}\,(\mathbf{y} - \phi\,\mathbf{x})^{\mathrm{T}}\,\Gamma^{(t)}(\mathbf{y} - \phi\,\mathbf{x}) + \lambda\, 1^{\mathrm{T}}\Lambda^{(t)}|\mathbf{x}|, \tag{3.34}$$

with

$$\Gamma^{(t)} = \begin{bmatrix} \alpha_1^{(t)} & 0 & \ldots & 0 \\ 0 & \alpha_2^{(t)} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \alpha_k^{(t)} \end{bmatrix}$$

and

$$\Lambda^{(t)} = \begin{bmatrix} \delta_1^{(t)} & 0 & \ldots & 0 \\ 0 & \delta_2^{(t)} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \delta_n^{(t)} \end{bmatrix}. \tag{3.35}$$

and where $|\mathbf{x}| = [|\mathbf{x}_1|, |\mathbf{x}_2|, \ldots, |\mathbf{x}_n|]^{\mathrm{T}}$.
The minimization of the function $L(\mathbf{x})$ is implemented, iteratively, as a succession of minimalizations of the function $Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)})$.

$$\begin{aligned} \hat{\mathbf{x}}^{(t+1)} &= \arg\min_{\mathbf{x}} Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)}) \\ \hat{\mathbf{x}}^{(t+1)} &= \arg\min_{\mathbf{x}} \frac{1}{2}\,(\mathbf{y} - \phi\,\mathbf{x})^{\mathrm{T}}\,\Gamma^{(t)}(\mathbf{y} - \phi\,\mathbf{x}) + \lambda\, 1^{\mathrm{T}}\Lambda^{(t)}|\mathbf{x}|. \end{aligned} \tag{3.36}$$

Greater detail in the deduction of the presented mathematical expressions can be found in (Jardim, 2008).

4. Majorization-Minimization Algorithm

The minimization of the function $Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)})$ (3.33) reflects an unconstrained convex optimization problem. In order to solve this problem with the algorithm GPSR-BB (Figueiredo *et al.*, 2007*b*), it is necessary to reformulate the optimization problem (3.36) as a quadratic program (Nocedal & Wright, 1999), which leads to

$$\min_{\mathbf{z}} \ \mathbf{c}^{\mathrm{T}}\mathbf{z} + \frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{B}\mathbf{z} \quad \text{subject to: } \ \mathbf{z} \geq 0, \tag{4.1}$$

where $\mathbf{z} = [\mathbf{u}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}}]^{\mathrm{T}}$ is a vector of unknown variables, with $\mathbf{u} = \max\{0, \mathbf{x}\}$ and $\mathbf{v} = \max\{0, -\mathbf{x}\}$, $\mathbf{c} = \lambda\, 1_{2n} + [-\mathbf{b}^{\mathrm{T}}, \mathbf{b}^{\mathrm{T}}]$ with $\mathbf{b} = \mathbf{A}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{A} = \phi^{\mathrm{T}}\Gamma$, and $\mathbf{B} = \begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{A} & -\mathbf{A}^{\mathrm{T}}\mathbf{A} \\ -\mathbf{A}^{\mathrm{T}}\mathbf{A} & \mathbf{A}^{\mathrm{T}}\mathbf{A} \end{bmatrix}$.

Considering the previously stated, the following algorithm was implemented in order to solve the optimization problem (4.1).

Step 0 (initialization): Given an initial estimate $\mathbf{z}^{(0)}$, set $t = 0$.

Step 1: Compute $\Gamma^{(t)}$ and $\tau^{(t)}$ according to (3.35) and $\tau_i = \lambda\, \Lambda_{ii}^{(t)}, \ \forall_{i=1,\dots,n}$, respectively.

Step 2: Execute GPSR-BB algorithm with entries the current estimate $\hat{\mathbf{x}}^{(t)}$, the $\tau^{(t)}$ vector, $\mathbf{A} = \Delta^{(t)}\phi$, and $\mathbf{y}_{\mathrm{m}} = \Delta^{(t)}\mathbf{y}$.

$$\hat{\mathbf{x}}^{(t+1)} = \text{GPSR-BB}(\mathbf{y}_{\mathrm{m}}, \mathbf{A}, \tau^{(t)}, \hat{\mathbf{x}}^{(t)}).$$

Step 3: Perform convergence test and terminate with approximated solution $\hat{\mathbf{x}}^{(t+1)}$; otherwise set $t = t + 1$ and return to **Step 1**.

4.1. Stopping Criterion

Initially we used the general stopping criterion

$$\frac{\|\hat{\mathbf{x}}^{(t+1)} - \hat{\mathbf{x}}^{(t)}\|_2}{\|\hat{\mathbf{x}}^{(t)}\|_2} \leq \varepsilon, \tag{4.2}$$

and it was found that it leads to good results. After that we choose a stopping criterion more directed to the problem to be solved, adopting the one it was used in the GPSR-BB algorithm, where the algorithm stops when the relative change in the number of nonzero components of the estimate falls below a given bound value.

4.2. Computational cost analysis

The number of iterations required to find an approximate solution, both in the outer cycle as in the inner one (**Step 2**), is not possible to accurately predict, since it depends (among other factors) on the quality of the initial estimate $\hat{\mathbf{x}}^{(0)}$. However, is possible to analyze the computational cost of each iteration of the proposed algorithm. For outer cycle, each iteration computational cost is essentially the inherent in the calculation of the matrix $\Gamma^{(t)}$, vectors $\tau^{(t)}$ and $\mathbf{y}_{\mathrm{m}}$. The computation of the matrix $\Gamma^{(t)}$ matrix, as well as the vector $\mathbf{y}_{\mathrm{m}}$, implies a matrix-vector product. To compute $\Gamma^{(t)}$ it is necessary to multiply the $(k \times n)$ matrix ϕ, by the vector of estimates $\hat{\mathbf{x}}^{(t)}$, of dimension n. This operation has a cost of $O(kn)$. To compute the vector $\mathbf{y}_{\mathrm{m}}$ is necessary to multiply the matrix $\Gamma^{(t)}$, of dimension $(k \times k)$, by the vector $\mathbf{y}$, of dimension k, which has a computational cost of $O(k)$. Calculate the vector $\tau^{(t)}$ implies vector-scalar product, which requires n floating-point operations.

The main computational cost per each cycle iteration of the GPSR-BB algorithm is a small number of inner products, scalar-vector multiplications, and vectors additions, each one of them requiring n or $2n$ floating-point operations, , plus a modest number of multiplications by $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}$. The algorithm proposed in this paper for the resolution of the optimization problem that results from the minimization of the function $L(\mathbf{x})$ (2.1), executes GPSR-BB algorithm, where the matrix $\mathbf{A}$ results from the product of the matrices ϕ and $\Gamma^{(t)}$. Given that ϕ is a $k \times n$ matrix, the computational cost of direct implementation of matrix-vector products by ϕ or ϕ^{T} is $O(kn)$. For $\Gamma^{(t)}$, $k \times k$ matrix, the computational cost of direct implementation of matrix-vector products is $O(k)$. If $\phi = \mathbf{RW}$ is a matrix of dimension $k \times n$ and $\mathbf{R}$ a $k \times d$ matrix, then $\mathbf{W}$ must be a $d \times n$ matrix. If $\mathbf{W}$ contains an orthogonal wavelet basis $(d = n)$, matrix-vector products involving $\mathbf{W}$ or $\mathbf{W}^{\mathrm{T}}$ can be implemented using fast wavelet tranform algorithms with $O(n)$ cost (Mallat, 1999), instead of the $O\left(n^2\right)$ cost of a direct matrix-vector product. Consequently, the cost of a product by ϕ or ϕ^{T} is O(n) plus that of multiplying by $\mathbf{R}$ or $\mathbf{R}^{\mathrm{T}}$ which, with a direct implementation, is $O(kn)$.

4.3. Convergence analysis

In order to analyze the convergence of the algorithm proposed in this paper, first will be analyzed the convergence of the GPSR-BB algorithm used in each iteration of the majorization-minimization algorithm, whose entries differ from those of the original algorithm of Figueiredo, being given by the equations defined above. Secondly will be analyzed the convergence of the iterative algorithm defined by the update (3.36).

As stated by Figueiredo in (Figueiredo *et al.*, 2007*b*) the convergence of the algorithm GPSR-BB used in this work can be derived from the analysis of Bertsekas (Bertsekas, 1999), but follows most directly from the results of Serafini, Zanghirati, and Zanni (Serafini *et al.*, 2005). In the algorithm proposed in this work we use the GPSR-BB algorithm with entries different from the ones defined by Figueiredo (Figueiredo *et al.*, 2007*b*). To summarize convergence properties of the GPSR-BB algorithm with the entries previously defined, we assume that termination occurs only when $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}$, which indicates that $\mathbf{x}^{(t)}$ is optimal.

Theorem 1: When $p = 1 \wedge q \in [1, 2]$ the sequence of iterates generated by the GPSR-BB algorithm with the entries $\mathbf{y}_{\mathrm{m}}$, $\mathbf{A}$, $\tau^{(t)}$ and the current estimate $\hat{\mathbf{x}}^{(t)}$ either terminates at a solution of (4.1) or else converges to a solution of (4.1) at an R-linear rate.

Proof. Theorem 2.1 of (Serafini *et al.*, 2005) can be used to demonstrate that all accumulation points of $\{\mathbf{x}^{(t)}\}$ are stationary points. Although, in (Serafini *et al.*, 2005), this result applies to an algorithm in which the steplength parameters $\alpha^{(t)}$ used in the gradient projection method are chosen by a different scheme, the only relevant requirement on these parameters in the proof of [(Serafini *et al.*, 2005), Theorem 2.1] is that they lie in the range $[\alpha_{\min}, \alpha_{\max}]$, as in the case of the GPSR-BB algorithm used in this work. When $p = 1 \wedge q \in [1, 2]$, the objective in (4.1) is convex and bounded below, we can apply [(Serafini *et al.*, 2005), Theorem 2.2] to deduce convergence to a solution of (4.1) at an R-linear rate. When $p < 1$, the objective function in (4.1) is nonconvex; thus it can not be guaranteed that the algorithm converges to a global optimum. Nevertheless, in practice, we have never observed any convergence problems: the results show that the proposed algorithm finds actually a minimum, which although may not be a global minimum corresponds to a good reconstruction of the signal. □

Theorem 2: For $Q(\mathbf{x}, \mathbf{x}')$ a continuous function in $(\mathbf{x}, \mathbf{x}')$ and L a strictly convex function, the MM iteration sequence $\hat{\mathbf{x}}^{(t)}$ converges to the global minimum of L.

Proof. Knowing that $\hat{\mathbf{x}}^{(t+1)} = \arg\min_{\mathbf{x}} Q(\mathbf{x}|\hat{\mathbf{x}}^{(t)})$, and recalling that the majorizer function Q verifies the conditions given by (3.6) and (3.7), we have

$$L(\hat{\mathbf{x}}^{(t+1)}) \leq Q(\hat{\mathbf{x}}^{(t+1)}|\hat{\mathbf{x}}^{(t)}) \leq Q(\hat{\mathbf{x}}^{(t)}|\hat{\mathbf{x}}^{(t)}) = L(\hat{\mathbf{x}}^{(t)}), \tag{4.3}$$

where the left hand inequality follows from the definition of Q and the right hand inequality from the definition of $\hat{\mathbf{x}}^{(t+1)}$. The sequence $L(\hat{\mathbf{x}}^{(t)})$, for $t = 1, 2, ...,$ is, therefore, nonincreasing. Under mild conditions, namely that $Q(\mathbf{x}, \mathbf{x}')$ is continuous in $(\mathbf{x}, \mathbf{x}')$, all limit points of the MM sequence $L(\hat{\mathbf{x}}^{(t)})$ are stationary points of L, and $L(\hat{\mathbf{x}}^{(t)})$ converges monotonically to $L^* = L(\mathbf{x}^*)$, for some stationary point $\mathbf{x}$. If, additionally, L is strictly convex, $\hat{\mathbf{x}}^{(t)}$ converges to the global minimum of L. Proofs of these properties are similar to those of similar properties of the EM algorithm (Huber & Ronchetti, 2009). □

5. Experiments

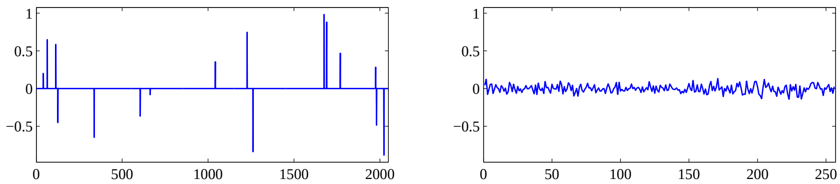
In this section are presented, analyzed and discussed the results obtained by the proposed algorithm in compressed sensing applications and in the reconstruction of sparse images or with sparse representations, where for each signal the algorithm is tested for different values of **p** and **q**. Parameter λ is hand-tuned for the best SNR improvement. For compressed sensing scenarios it is adjusted according to the expression $\lambda = 0.1\,\|\phi^T\mathbf{y}\|_\infty$ (as suggested by Fuchs in (Fuchs, 2004)).

All the experiments reported in this section were obtained with MATLAB (MATLAB 7.0 R14) implementations of the algorithm described above. The computing platform is a standard personal computer with Intel(R) Core(TM) i7 CPU, 8 GB of RAM, and running Windows 7 operating system.

5.1. Compressed Sensing

We first consider a typical compressed sensing (CS) scenario, where the goal is to reconstruct a length-n sparse signal (in the canonical basis, thus $\mathbf{W} = \mathbf{I}$ and $d = n$) from k observations, where

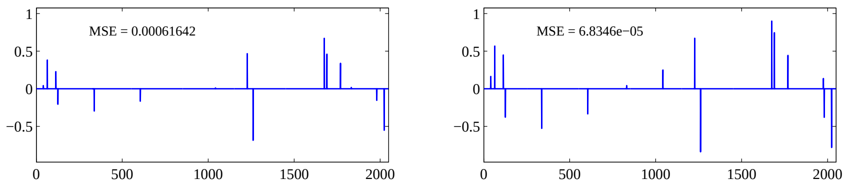
$k \leq n$. The rows of the $k \times n$ observation matrix $\mathbf{R}$ are unit-norm random vectors (of Gaussian components) in $\Re^n$. Notice that, since $k \leq n$, the system $\mathbf{Rx} = \mathbf{y}$ is undetermined. In the first example, we take $n = 2^{11} = 2048$, $k = 2^8 = 256$, and generate y by adding Laplacian noise (probability density function $f(x) = a\ e^{-\frac{|x|}{a}}$) with parameter $a = 0.01$) to $\mathbf{Rx}$ (figure (2(b))). The original sparse signal $\mathbf{x}$ is generated by a mixture of a uniform distribution on $[1, 1]$ and a point mass at zero, with probabilities 0.01 and 0.99, respectively. As we can see in 2(a), the original signal is indeed sparse.



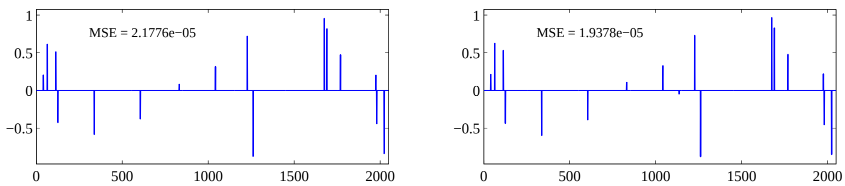
(a) Original signal. (b) Observed signal.

Figure 2. Original and observed signal.

For the described data set we have as initial estimate for ℓ_q - ℓ_p algorithm the signal $\hat{\mathbf{x}}^{(0)} = \phi^T \mathbf{y}$. The estimates obtained by solving (2.1) using the proposed algorithm for $q = 1 \wedge p \in]0, 1]$ are shown in figure (3), and it can be verified that they are very close to the original signal.



(a) $q = 1 \wedge p = 1$. (b) $q = 1 \wedge p = 0.75$.



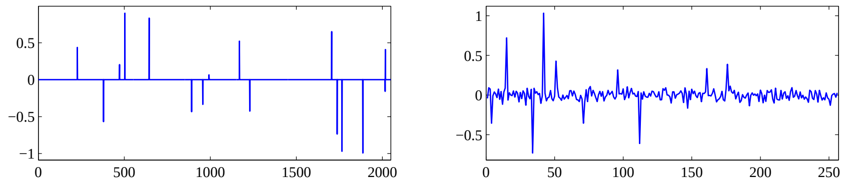
(c) $q = 1 \wedge p = 0.5$. (d) $q = 1 \wedge p = 0.25$.

Figure 3. Estimated signal by (2.1) minimization for $q = 1 \wedge p \in]0, 1]$

From the results obtained, and presented in figure (3) it is possible to observe that the approximation mean squared error[b] (MSE) for $q = 1 \wedge p = 1$ is about 10 times greater than that obtained for $p \leq 1$, meeting the expected results.

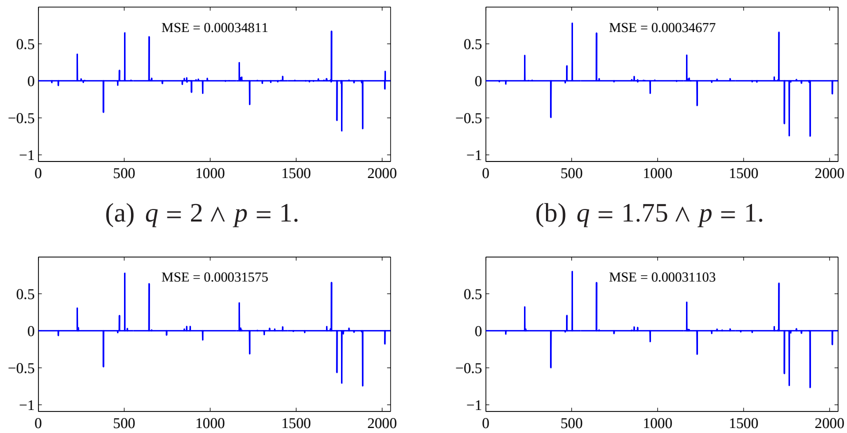
[b]**MSE** $= (1/n)\ ||\hat{\mathbf{x}} - \mathbf{x}||_2^2$, where $\hat{\mathbf{x}}$ is an estimate of $\mathbf{x}$.

As referred above, the generalization of the data term from a ℓ_2 norm to a ℓ_q norm, gives to the approximation criterion statistical strength features when $q < 2$. This can be verified in the second experiment, where we take $n = 2^{11} = 2048$, $k = 2^8 = 256$, and generate y by adding to **Rx** impulsive noise. Considering the original signal represented in figure (4(a)) with 16 nonzero components, we can see in figure (5) that the MSE of the approximation decreases with the value of q, taken as constant the value of $p = 1$.



(a) Original signal. (b) Observed signal.

Figure 4. Original and observed signal.



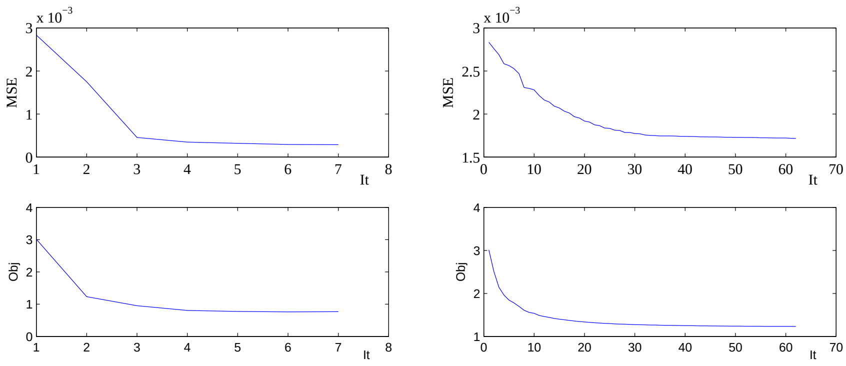
(a) $q = 2 \wedge p = 1$. (b) $q = 1.75 \wedge p = 1$.

(c) $q = 1.5 \wedge p = 1$. (d) $q = 1.25 \wedge p = 1$.

Figure 5. Estimated signal by (2.1) minimization for $q \in]1, 2] \wedge p = 1$

Note, for example, that the best result achieved by ℓ_q - ℓ_p algorithm for $q = 2$ (MSE = 3.4811×10^{-4}, 59 nonzero components) is not so good as the obtained for $q = 1.25$ (MSE = 3.1103×10^{-4}, 35 nonzero components). Although a significant improvement in quantitative terms (the values of the approximations MSE for different values of q are not too different) doesn't occur, in this case, we can verify that, as the value of q decreases, the approach presents qualitative improvements (lower number of spurious observations in the final estimate). Figures (6(a)) and (6(b)) show the evolution of the MSE and objective function, for outer and inner loop respectively, against iteration number of the proposed algorithm, when $q = 1.5 \wedge p = 1$.

Analyzing the graphics of figure (6) we can observe that the values of the approximation MSE and the objective function decreases in each iteration of the $\ell_q - \ell_p$ algorithm. In fact, starting



(a) Algorithm outer loop (b) Algorithm inner loop

Figure 6. MSE and Objective function *vs.* Iteration number

from a initial upper bound function, built on the initial estimate, it is observed that the proposed algorithm builds at each iteration a new one, on the basis of the obtained solution for the previous upper bound function, which solution is closest to the original signal.

Next experiment shows the performance of the implemented algorithm in a typical compressed sensing problem, where the goal is to reconstruct a signal from k projections (with $k = 2^{10} = 1024$). The observation matrix $\mathbf{R}$, of dimension $k \times n$, is a matrix of random projection vectors. The two-dimensional original signal, which is represented in figure (7) has a sparse wavelet transform, and in this example the columns of the representation matrix $\mathbf{W}$ form an orthogonal wavelet basis (daubechies-2 (Haar)). The original signal is an image of piecewise smooth filtered white noise of dimension $n' \times n'$, with $n' = 2^6$, *i. e.* $n = 2^{12}$ (figure (7)). The observed signal is obtained by adding white Gaussian noise with standard deviation 0.001 to $\mathbf{Rx}$. Figures (8(a)) - (8(c)) shows the results of ℓ_q - ℓ_p algorithm, taken as initial estimate $\hat{\mathbf{x}}^{(0)} = 0$. By the results we can see that the proposed algorithm produces, from k projections corrupted by random white Gaussian noise best approximations (lower MSE) for $p < 1$.

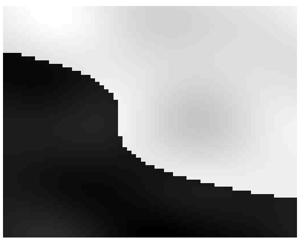
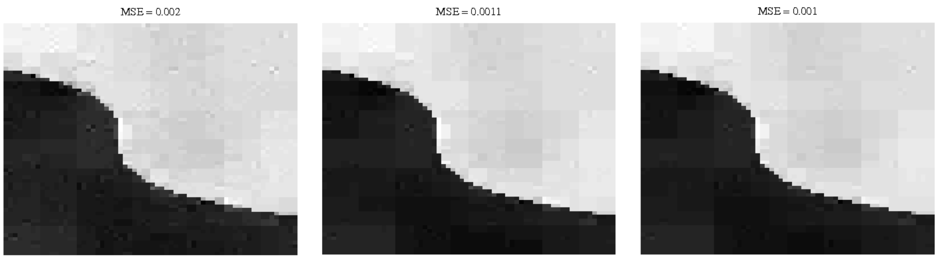


Figure 7. Original signal.

5.2. *Image Restoration*

The following two experiments show the performance of the proposed algorithm in real images restoration, where the columns of representation matrix $\mathbf{W}$ form an orthogonal wavelet basis



(a) $q = 2 \wedge p = 1$. (b) $q = 2 \wedge p = 0.4$. (c) $q = 2 \wedge p = 0.2$.

Figure 8. Estimated signal by (2.1) minimization, for $q = 2 \wedge p \in]0,\ 1]$.

(daubechies-2 (Haar)). The observation matrix $\mathbf{R}$, of dimension $k \times k$, is a Toeplitz blocks matrix representing 2D convolutions. In the first of these experiments we have as original signal the well-known Camera-man image (figure (9(a))). The observed image is obtained convolving the original image with a uniform blur filter of size 9×9, and then adding to the blurred image ($\mathbf{Rx}$) white Gaussian noise with standard deviation 0.005 (figure (9(b))). Taken as initial estimate $\hat{\mathbf{x}}^{(0)} = \phi^{T}\mathbf{y}$, figures (10(a)) - (10(d)) shows the algorithm performance for $q = 2$ and $p \in]0,\ 1]$. For $q = 2 \wedge p = 0.5$ the initial value of the objective function is 2.0834×10^{6} and the final one is 3.146×10^{4}.

A typical statistical model for image wavelet coefficients is the Generalized Gaussian Density (Moulin & Liu, 1999), given by $p(\theta) = \exp\left\{-\frac{|\boldsymbol{\theta}|^{p}}{\alpha}\right\}$, where $\boldsymbol{\theta}$ represents the wavelet coefficients vector. The graph shown in figure (11) represents the evolution of MSE with p, where we can see that the best approximation occurs for $p = 0.5$. This is consistent with Moulin's statment (Moulin & Liu, 1999) that good image models based on wavelets are obtained doing $p \simeq 0.7$.

(a) Original image. (b) Observed image.

Figure 9. Original and observed real image.

In the last reported experiment the original image is the also well-known Lenna image (figure (12(a))). The observed image is obtained convolving the original image with a uniform blur filter of size 9×9, and then adding to the blurred image ($\mathbf{Rx}$) impulsive noise (figure (12(b))).

Figures (13(a)) - (13(d)) shows the performance of the implemented algorithm for $q \in]1,\ 2]$ and $p = 1$. Again we can observe that best results are achieved for values of p below 1. For this experiment, and for $q = 1.25 \wedge p = 1$, we have that the initial value of the objective function is 3.7745×10^{6} and the final one is 1.689×10^{5}.

(a) $q = 2 \wedge p = 1$.

(b) $q = 2 \wedge p = 0.8$.

(c) $q = 2 \wedge p = 0.6$.

(d) $q = 2 \wedge p = 0.5$.

Figure 10. Estimated signal by (2.1) minimization, for $q = 2 \wedge p \in]0,\ 1]$.

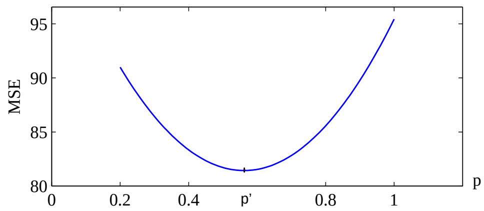


Figure 11. Approximation MSE evolution with p.

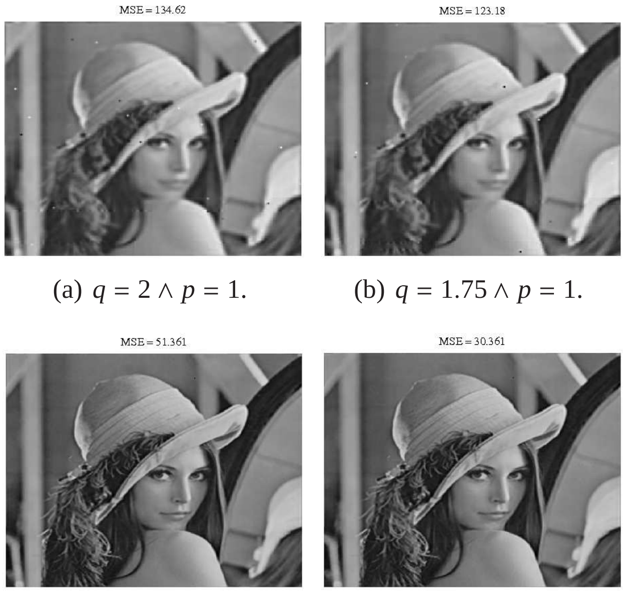
(a) Original image.

(b) Observed image.

Figure 12. Original and observed real image.

6. Conclusions

In this paper was proposed a majorization-minimization class algorithm to address ℓ_q-ℓ_p optimization problems, where $p \in]0,\ 1] \wedge q \in [1,\ 2]$, of which ℓ_2-ℓ_1 is an instance. The proposed algorithm was tested on scenarios of compressed sensing and image reconstruction, and, in both



(a) $q = 2 \wedge p = 1$. (b) $q = 1.75 \wedge p = 1$.

(c) $q = 1.5 \wedge p = 1$. (d) $q = 1.25 \wedge p = 1$.

Figure 13. Estimated signal by (2.1) minimization, for $q \in]1, 2]$ and $p = 1$.

cases, experiments were performed for data corrupted by uniform and impulsive noise. As mentioned, the generalization of the coefficients term to be estimated, from a ℓ_1 to a ℓ_p norm, with $p < 1$, aims to move towards the solution of the original combinatorial optimization problem, whose complexity prevents the calculation of a global solution in polynomial time. Although no one can guarantee the convergence of the algorithm to a global minimum, whereas for $p < 1$ the problem is no longer convex,the analysis of the results shows that the proposed algorithm provides as solution of the optimization problem a minimum corresponding to a signal reconstruction better than the one obtained by making $p = 1$ (lowest approximation MSE).

While in the experiments with nonnatural sparse signals we verify that approximation MSE, relatively to the original signal, decreases with the value of p, the same is not true for natural images, where we verify that there is an optimal value of $p < 1$, for which we obtain the best possible approximation, *i.e.*, that which corresponds the smallest MSE. This is justified against the model used for wavelet coefficients (GGD - Generalized Gaussian Density), which is function of the parameter p.

In the presence of outliers the proposed algorithm was tested taking p constant and equal to 1, and ranging q in $]1, 2]$. Both in compressed sensing applications of unidimensional signals, as in natural images restoration applications we can verify that the approximation MSE decreases with the value of parameter q. From the analysis of the results we can still observe that, as the value of q decreases the amount of outliers of the optimization problem solution also decreases. Hence it can be concluded that the use of ℓ_q norms, with $q < 2$, in data term, gives to the approximation criterion statistical robustness characteristics.

References

Bertsekas, D. P. (1999). *Nonlinear Programming*. 2nd edition ed.. Athena Scientific.

Candes, E. and T. Tao (2005). Decoding by linear programming. *IEEE Transactions on Information Theory* **51**(12), 4203–4215.

Casas, E., C. Clason and K. Kunisch (2012). Approximation of elliptic control problems in measure spaces with sparse solutions. *SIAM Journal on Control and Optimization* **50(4)**, 1735–1752.

Chen, S., D. L. Donoho and M. A. Saunders (2001). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific and Statistical Computing, Review* **43**, 129–159.

Claerbout, J. and F. Muir (1973). Robust modelling of erratic data. *Geophysics* **38**, 826–844.

Davis, G., S. Mallat and M. Avellaneda (1997). Adaptive greedy approximation. *Journal Constructive Approximations* **13**(1), 57–98.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory* **52**, 1289–1306.

Efron, B., T. Hastie, I. Johnstone and R. Tibshirani (2004). Least angle regression. *The Annals of Statistics* **32**(2), 407–499.

Figueiredo, M. A. T. and R. Nowak (2003). An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing* **12**, 906–916.

Figueiredo, M. and R. Nowak (2005). A bound optimization approach to wavelet-based image deconvolution. In: *IEEE International Conference on Image Processing - ICIP'2005*.

Figueiredo, M., J. Bioucas-Dias and R. Nowak (2007*a*). Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing: Special Issue on Convex Optimization Methods for Signal Processing* **16**(12), 2992–3004.

Figueiredo, M., R. Nowak and S. Wright (2007*b*). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing* **1**(4), 586–598.

Fuchs, J.-J. (2004). On sparse representations in arbitrary bases. *IEEE Transactions on Information Theory* **50**, 1341–1344.

Herzog, R., G. Stadler and G. Wachsmuth (2012). Directional sparsity in optimal control of partial differential equations. *SIAM Journal on Control and Optimization* **50(2)**, 943–963.

Huber, P. J. and E. M. Ronchetti (2009). *Robust Statistics*. 2nd edition ed.. Wiley.

Hunter, D. and K. Lange (2004). A tutorial on mm algorithms. *The American Statistician* **58**, 30–37.

Jardim, S. (2008). Algoritmos para Representacao Esparsa e Robusta de Sinais. PhD thesis. Instituto Superior Tecnico, Universidade Tecnica de Lisboa. In Portuguese.

Malioutov, D., M. Cetin and A. Willsky (2005). Homotopy continuation for sparse signal representation. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 5. Philadelphia. pp. 733–736.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. 2nd edition ed.. Academic Press.

Moulin, P. and J. Liu (1999). Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors. *IEEE Transaction on Information Theory* **45**, 909–919.

Natarajan, B. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing* **24**(2), 227–234.

Nocedal, J. and S. Wright (1999). *Numerical Optimization*. Springer-Verlag. New York.

Serafini, T., G. Zanghirati and L. Zanni (2005). Gradient projection methods for quadratic programs and applications in training support vector machines. *Optimization Methods and Software*.

Stadler, G. (2009). Elliptic optimal control problems with l1-control cost and applications for the placement of control devices. *Computational Optimization and Applications* **44**(2), 159–181.

Turlach, B. (2005). On algorithms for solving least squares problems under l1 penalty or an l1 constraint. In: *Proceedings of the American Statistical Association; Statistical Computing Section*. Alexandria. pp. 2572–2577.

Turlach, B., W. N. Venables and S. J. Wright (2005). Simultaneous variable selection. *Technometrics* **27**, 349–363.

Wright, S. J., R. D. Nowak and M. A. T. Figueiredo (2009). Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing* **57**, 2479–2493.

Zarzer, C. A. (2009). On tikhonov regularization with non-convex sparsity constraints. *Inverse Problems* **25:025006**, 13 pp.