Canad. Math. Bull. Vol. 63 (4), 2020 pp. 726–737
http://dx.doi.org/10.4153/S0008439519000730

Non-convex Optimization via Strongly Convex Majorization-minimization

Azita Mayeli

Abstract. In this paper, we introduce a class of nonsmooth nonconvex optimization problems, and we propose to use a local iterative minimization-majorization (MM) algorithm to find an optimal solution for the optimization problem. The cost functions in our optimization problems are an extension of convex functions with MC separable penalty, which were previously introduced by Ivan Selesnick. These functions are not convex; therefore, convex optimization methods cannot be applied here to prove the existence of optimal minimum point for these functions. For our purpose, we use convex analysis tools to first construct a class of convex majorizers, which approximate the value of non-convex cost function locally, then use the MM algorithm to prove the existence of local minimum. The convergence of the algorithm is guaranteed when the iterative points $x^{(k)}$ are obtained in a ball centred at $x^{(k-1)}$ with small radius. We prove that the algorithm converges to a stationary point (local minimum) of cost function when the surregators are strongly convex.

1 Introduction

Consider the following optimization problem

$$\min_{x\in\mathcal{C}} F(x), \tag{1.1}$$

where $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^N$ and $F\colon \mathbb{R}^N \to \mathbb{R}$ is a real valued objective or cost function. In general, F is continuous but not convex nor smooth. Most optimization problems rely heavily on convexity condition of the function F, and the lack of convexity for F usually makes it an NP hard problem to find a global minimum point for the optimization problem (1.1). In particular, the convexity condition is useful in some practical problems such as in image reconstruction and sparse recovery [20, 21]. In the absence of the convexity condition, *majorization-minimization* (MM) algorithm has been proved to be a useful tool in finding local minimization vectors or signals. This algorithm is an iterative algorithm, and it converts a difficult optimization problem into a simple one, as we will demonstrate in some of these cases in this paper.

Received by the editors June 24, 2019; revised November 25, 2019.
Published online on Cambridge Core December 10, 2019.
AMS subject classification: **65R32**, 90C26.
Keywords: Cost function, local majorizer and minimizer, surregator, Moreau envelope, infimal convolution, convex function, stationary point.

The goal of this paper is to solve the following class of problems using an iterative algorithm: for given $y \in \mathbb{R}^M$,

$$\underset{x \in \mathbb{R}^N}{\arg\min}\, F(x), \tag{1.2}$$

where $F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda(\|x\|_1 - f_\alpha(x)), \forall x \in \mathbb{R}^N$.

Here, f_α is the Moreau envelope of a convex function as defined in (2.2), and $\alpha > 0, \lambda > 0$ are constants and predetermined. The matrix $A \in \mathbb{R}^{M\times N}$ is a low rank wide matrix (*e.g.*, a finite frame or wavelet). In our setting, the penalty term is a non-convex function and is given by

$$\psi_\lambda(x) = \lambda(\|x\|_1 - f_\alpha(x)). \tag{1.3}$$

The optimization problem (1.2) is a nonconvex nonsmooth optimization problem subject to the penalty function ψ_λ. The cost function F given by (1.2) is in general nonconvex nonsmooth. However, the convexity can hold under some conditions depending on the choice of A, λ, and α. Note that the main idea of using such nonconvex penalty functions is to promote the sparsity of the solutions in (1.2). A non-convex penalty can induce a nonconvex cost function, thus unnecessary suboptimal local minimizers for the cost function. The main goal of this paper is twofold. First, we introduce a class of functions that majorize the cost function locally. Then we use these majorizers (surrogaters) in an MM algorithm to solve the optimization problem (1.2) and prove that the iteration points convergence to the stationary point of the objective function under some sufficient condition. Before we explain the main contributions of the current work in details, let us first recall some known and special cases of (1.2).

Special cases When $\lambda = 0$, the problem is alternately referred to as a minimizer of the residual sum of squared errors (RSS). The solution for minimization can be obtained by the least square method. In this case, the minimization is a continuously differentiable, unconstrained, convex optimization problem. For a solution of this case, see *e.g.*, ([13]). When f_α is a constant function (e.g. when $\alpha = 0$), the problem turns into the classical ℓ_1 regularizer case. This case, among the cases with convex regularizer (or penalty), is more effective in inducing sparse solutions for (1.1) and (1.2) ([4]). However, the ℓ_1 regularizer underestimates the high amplitude components of the solution. The problem with an ℓ_1 penalty is known as the *Least Absolute Selection and Shrinkage Operator* (LASSO) ([23]) and *Basis Pursuit Denoising* ([7]). Several methods have been introduced in [7, 23] for optimizing the problem. When $f(x) = \|x\|_1$, the Moreau envelope f_α is the well-known Huber function. The Huber function and its general form as regulizers of sparse recovery problems have been treated in [19], and it has been proved that with these regularizers, using proximal algorithms, problem (1.2) has an optimal solution (global minimum) provided that F is convex. In this case, the penalty term (1.3) is called *MC penalty*.

Main contribution The first contribution of this paper is to construct a class of convex functions that majorize (surrogate) the cost function F (1.2) locally. We obtain these functions by constructing local minimizers for the penalty term ψ_λ. The local majorizers are tangent to the cost function only at one point and each has a global

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

minimum. The existence of a global minimum for the majorizers is obtained by convexity of majorizers, which we also study here.

The second contribution of this paper is to use the MM algorithm to find a sequence of iteration points that converges to the local minimum of the cost function (1.2). In this algorithm, each iteration point $x^{(k)}$ is obtained by local minimization of surregator function $F^M(\cdot, x^{(k-1)})$ in some small neighbourhood of $x^{(k-1)}$. We prove that the sequence $\{x^{(k)}\}_k$ has an accumulation point and is a stationary point for the cost function F in (1.2), provided that the majorizers are a-strongly convex.

Outline The paper is organized as follows. After introducing some notations and preliminaries in Section 2, in Section 3 we introduce a class of minimizer functions for the penalty term (1.3) to obtain majorizers for the cost function F (1.2). In this section, we also study sufficient conditions for the majorizers to be convex. These results are collected in Lemma 3.1 and Theorem 3.2, respectively. In Section 4, we propose to use the iterative MM algorithm to obtain a stationary point (local minimum) of F. These results are collected in Proposition 4.1, Theorem 4.4, and Corollary 4.5.

1.1 Related Work

The current paper proposes the use of majorization-minimization (MM) algorithm to solve the class of nonconvex nonsmooth optimization problems of type (1.2). The MM approach has been used, for example, in [10, 14, 16] for solving some nonconvex optimization problems different from those we consider here. There are other types of methods that have been proved effective in solving nonconvex problems, *e.g.*, iteratively reweighted least squares (IRLS) method ([9]) and iteratively reweighted ℓ_1 (IRL1) ([5]). For a list of other methods including gradient descent method to find local minimum of a function, we refer the reader to [16] and the reference therein.

Our approach to solving a non-convex optimization problem involves the minimization of a cost function defined in terms of the penalty function ψ_λ (1.3) with the regularization parameter λ. The minimization of a cost function defined in terms of the ℓ_1 norm, $\lambda\| \, \|_1$, can be obtained via convex optimization techniques. This has been considered in a sequence of works, *e.g.*, [3, 8, 22]. Indeed, since its early application in geophysics, the ℓ_1 norm and sparsity have become important tools in signal processing [1]. The most well-known nonconvex regularizer is the ℓ_p pseudo-norm $(0 \le p < 1)$ [24, 26]. For example, when $p = \frac{1}{2}$, an iterative technique, called the *iterative half thresholding algorithm*, is used to obtain a local minimizer of the $\ell_{\frac{1}{2}}$ regularizer.

An analogy to soft thresholding for the ℓ_1-penalty is the non-convex log-thresholding $\lambda \sum_i \log(\delta + |x_i|)$, and the proximal splitting step of the algorithm has a closed form solution ([17]). Their technique is based on a direct link of reweighted ℓ_1-penalties (IRL1) to the concave log-regularizer for sparsity. The minimization of an objective function with the penalty term as the difference of ℓ_1 and ℓ_2 norms, $\lambda(\|\cdot\|_1 - \|\cdot\|_2)$, and as a non-convex and Lipschitz continuous metric for solving constrained and unconstrained compressed sensing problems, has been studied in [25]. The solution of the optimization problem with this penalty has been obtained via the difference of convex algorithm (DCA). Other regularizers have also been advocated in [6, 11, 12].

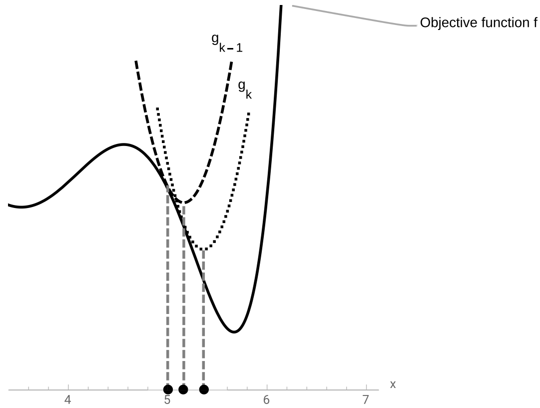

Figure 1: The MM algorithm procedure. The nodes "• • •" on the horizontal line represent the iteration points $x^{(k-1)}$, $x^{(k)}$, and $x^{(k+1)}$.

2 Preliminaries and Notation

For any vector $x \in \mathbb{R}^N$, the ℓ_1 and ℓ_2 norms of x are defined by $\|x\|_1 = \sum_i |x_i|$ and $\|x\|_2^2 = \sum_i |x_i|^2$, respectively. We denote a matrix of dimension $M \times N$ by $A \in \mathbb{R}^{M\times N}$. We say it is *positive semidefinite*, denoted by $A \geq 0$, if for all $x \in \mathbb{R}^N$, $\langle Ax, x\rangle \geq 0$. Here, $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. Positive definiteness is also equivalent to saying that all eigenvalues of A are non-negative. We say two functions f and h are tangent at the point w when f and h both have directional derivatives at w and for any direction $d \in \mathbb{R}^N$ with small $\|d\|_2$, $\nabla f(w; d) = \nabla h(w; d)$.

Local majorizers and minimizers Given a fixed point $w \in \mathbb{R}^N$, a function $g(\cdot, w)\colon \mathbb{R}^N \to \mathbb{R}$ is called a *local majorizer* of function $f\colon \mathbb{R}^N \to \mathbb{R}$ at w if the following conditions hold:

$$\begin{aligned} f(x) &\leq g(x, w) \quad \forall x \in \mathbb{R}^N; \\ f(x) &= g(x, w) \quad \text{if and only if } x = w. \end{aligned}$$

From the point of view of geometry, a majorizer means that the surface obtained by the map $x \mapsto g(x, w)$ lies above the surface generated by $x \mapsto f(x)$, and these two surfaces are touching (have a tangent point) only at $x = w$.

We say a function $g(\cdot, w)\colon \mathbb{R}^N \to \mathbb{R}$ minorizes the function f at w when $-g(\cdot, w)\colon \mathbb{R}^N \to \mathbb{R}$ majorizes $-f$ at w.

The Iterative MM algorithm In the MM algorithm, we choose the majorizer $g_{k-1} := g(\cdot, x^{(k-1)})$ tangent to the objective function at $x^{(k-1)}$ and minimize it on a convex set $\mathcal{D}$ to obtain the next iteration point $x^{(k)}$. That is, $x^{(k)} := \arg\min_{x\in\mathcal{D}} g(x, x^{(k-1)})$, provided that $x^{(k)}$ exists. Then we define $g_k := g(\cdot, x^{(k)})$ (see Figure 1).

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

When the minimization points $x^{(k)}$ exist, the following descending property holds:

$$f(x^{(k)}) \leq g_{k-1}(x^{(k)}) = g(x^{(k)}, x^{(k-1)}) \leq g_{k-1}(x^{(k-1)}) = g(x^{(k-1)}, x^{(k-1)}) = f(x^{(k-1)}). \tag{2.1}$$

One of the significant properties of the MM algorithm is its stability due to the descending property of the objective function f (2.1). If an objective function is strictly convex, then the MM algorithm will converge to the unique optimal point (global minimum), assuming that it exists. In the absence of convexity, all stationary points are isolated; then the MM algorithm will converge to one of them. For a complete philosophy of the MM algorithm, we refer the reader to [14, 15] for example.

Moreau envelope For a function $\tilde{f}\colon \mathbb{R}^N \to \mathbb{R}$ and $\alpha > 0$, the Moreau envelope of $\tilde{f}$ is denoted by f_α and is defined by infimal convolution:

$$f_\alpha(x) := \Big(\tilde{f} \,\square\, \frac{\alpha}{2}\|\cdot\|_2^2\Big)(x) = \inf_{v\in\mathbb{R}^N} \Big\{\tilde{f}(v) + \frac{\alpha}{2}\|v - x\|_2^2\Big\}, \quad \forall x \in \mathbb{R}^N. \tag{2.2}$$

The function f_α is convex when $\tilde{f}$ is convex and it is the infimal convolution of the function $\tilde{f}$ and the map $x \mapsto \frac{\alpha}{2}\|x\|_2^2$. For example, when $\tilde{f}(x) = \|x\|_1$, the Moreau envelope $f_\alpha\colon \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ is a well-known (generalized) Hubber function. For the definition of infimal convolution and its other properties see, *e.g.*, [2].

Let $y \in \mathbb{R}^M$ be an observed vector data and let $A \in \mathbb{R}^{M\times N}$ be a matrix, which is usually a wide low rank matrix. The following result for the cost function F (1.2) with penalty term $\psi_\lambda(x) = \lambda(\|x\|_1 - f_\alpha(x))$ is a mild improvement of [19, Theorem 1]. In [19], ψ_λ is the MC penalty and the Moreau envelope f_α is the generalized Huber function.

Theorem 2.1 *The function F is (strictly) convex if*

$$A^T A - \lambda\alpha I \geq 0 \quad \textit{(convexity condition).} \tag{2.3}$$

For strictly convex functions, the inequality $\geq$ is replaced by $>$ 0. Here, I is the identity matrix, and λ and α are constants.

This theorem can be proved using a similar technique to that used to prove [19, Theorem 1]. Note that the condition (2.3) ensures the uniqueness of the minimizer of the cost function F.

The sufficient convexity condition (2.3) indicates that all eigenvalues of matrix $A^T A$ must be at least $\lambda\alpha$. In the absence of convexity, the function F is the sum of one concave function and one convex function, and it can have many local minimums. In this case, one needs an approach to prove the existence of a global minimum or global optimum point for F. This paper proposes the use of the MM algorithm technique for this purpose when F is nonconvex.

To reach our goal and prove the existence of a local minimizer for nonconvex objective (or cost) function F (1.2), we first construct local minimizers for the Moreau envelope f_α and then use them to obtain local majorizers for F. Our technique follows.

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

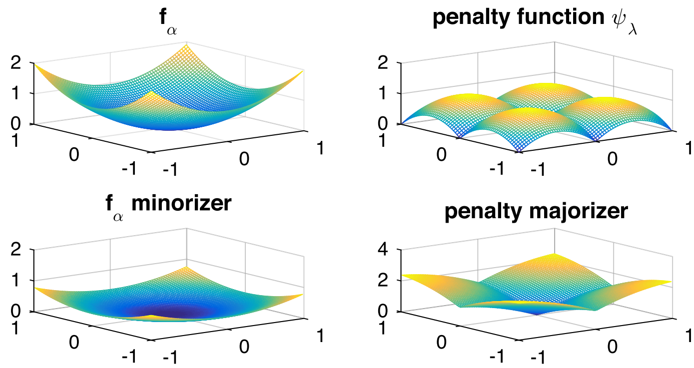

Figure 2: In this example, we put $\alpha = 1$ and $\lambda = 2$, and f_α is the Huber function. The penalty function ψ_λ is defined as in (1.3).

Let $\gamma_m > 0$ be a constant to be determined later. For any $w \in \mathbb{R}^N$, let

(2.4) $$f_\alpha^m(x,w) := f_\alpha(x) - \gamma_m \|x - w\|_2^2.$$

We define $F^M(\cdot, w)$ by replacing f_α by $f_\alpha^m(x,w)$ in the definition of the function F (1.2) as follows:

(2.5) $$F^M(x,w) := \frac{1}{2}\|y - Ax\|_2^2 + \lambda\big(\|x\|_1 - f_\alpha^m(x,w)\big).$$

It is obvious that $F^M(\cdot, w)$ majorizes F; *i.e.*, for all $x \in \mathbb{R}^N$, $F^M(x,w) \ge F(x)$. In Theorem 3.2, we prove that the surface generated by the function $F^M(\cdot, w)$ is lying about the surface generated by the function F and they touch only at one point, $x = w$. For an illustration of an infimal convolution f_α, a penalizer ψ_λ, and their minorizer and majorizer functions, respectively, see Figure 2.

Remark 2.2 In the same fashion, one can define minorizers $F^m(x,w)$ for F. For this, let $\gamma_M > 0$ and define $f_\alpha^M(x,w) := f_\alpha(x) + \gamma_M\|x - w\|_2^2$. Then $f_\alpha^M(x,w) \ge f_\alpha(x)$ for all x. Define

$$F^m(x,w) := \frac{1}{2}\|y - Ax\|_2^2 + \lambda\big(\|x\|_1 - f_\alpha^M(x,w)\big).$$

With similar proof techniques for majorizers in the rest of this paper, one can obtain local minorizers for the cost function F with tangential point at w. Minorizers are a useful tool in finding local maximums of an optimization problem.

3 Construction of a Local Majorizer for Cost Function

Our first result in this section proves the existence of local minorizers for the Moreau envelope function f_α, followed by the construction of majorizers for the cost function F.

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

Lemma 3.1 (Minorizer of f_α) *Fix $w \in \mathbb{R}^n$ and define $f_\alpha^m(\cdot, w)$ as in (2.4). Then $f_\alpha^m(\cdot, w)$ is a minorizer for f_α, and for any direction $d \in \mathbb{R}^N$ with $\|d\|_2$ small, we have*

$$\nabla f_\alpha(w; d) = \nabla f_\alpha^m(w; d, w). \tag{3.1}$$

Proof The proof of the local miniorizers for f_α is obtained directly from the definition of $f^m(\cdot, w)$. To prove (3.1), let $d \in \mathbb{R}^N$ with $\|d\|_2$ small. Then

$$\begin{aligned}
\nabla f_\alpha^m(w; d, w) & \\
&= \liminf_{\theta \to 0^+} \frac{f_\alpha^m(w + \theta d, w) - f_\alpha^m(w, w)}{\theta} \\
&= \liminf_{\theta \to 0^+} \frac{f_\alpha^m(w + \theta d, w) - f_\alpha(w)}{\theta} \quad (f_\alpha^m(w, w) = f_\alpha(w)) \\
&= \liminf_{\theta \to 0^+} \frac{\big(f_\alpha(w + \theta d) - \gamma_m \|\theta d\|_2^2\big) - f_\alpha(w)}{\theta} \quad (\text{by the definition of } f_\alpha^m) \\
&= \liminf_{\theta \to 0^+} \frac{\big(f_\alpha(w + \theta d) - f_\alpha(w)\big) - \gamma_m \|\theta d\|_2^2}{\theta} \\
&= \liminf_{\theta \to 0^+} \frac{\big(f_\alpha(w + \theta d) - f_\alpha(w)\big)}{\theta} - \liminf_{\theta \to 0^+} \gamma_m \theta \|d\|_2^2 \\
&= \liminf_{\theta \to 0^+} \frac{\big(f_\alpha(w + \theta d) - f_\alpha(w)\big)}{\theta} \\
&= \nabla f_\alpha(w; d).
\end{aligned}$$

This completes the proof of the theorem. ∎

Our next result illustrates that the local minorizers of the Moreau envelope function f_α induce local majorizers for F.

Theorem 3.2 *The function $F^M(\cdot, w)$ (2.5) is local majorizer for the cost function F at w, and we have the following.*

(i) *$\nabla F(w; d) = \nabla F^m(w; d, w)$ for all d with $\|d\|_2$ sufficiently small.*

(ii) *$F^M(\cdot, w)$ is convex if*

$$A^T A + \lambda(2\gamma_m - \alpha) I \geq 0. \tag{3.2}$$

The convexity is strict if $A^T A + \lambda(2\gamma_m - \alpha) I > 0$.

Proof By Proposition 3.1, it is immediate that the function $F^M(\cdot, w)$ is a local majorizer for F. Item (i) also holds by the equality in (3.1). To prove item (ii), we will adapt an approach used to prove [19, Theorem 1].

Notice that the discrepancy with respect to the data in the surregator function $F^M(\cdot, w)$ can be written as

$$F^M(x, w) = x^T\Big(\frac{1}{2} A^T A + \lambda\Big(\gamma_m - \frac{\alpha}{2}\Big) I\Big) x + \lambda \|x\|_1 + \max_{v \in \mathbb{R}^N} g(v, x, w). \tag{3.3}$$

Notice the function $Q(x) := \max_{v \in \mathbb{R}^N} g(v, x, w)$ is not affine, although for any fixed point (v, w), the map $x \to g(v, x, w)$ is affine (*i.e.*, linear). However, the convexity of Q can be obtained as a result of [2, Proposition 8.14], since Q is the pointwise

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

maximum of convex functions. Therefore by (3.3), $F^M(\cdot, w)$ is convex when the quadratic part is convex. This means that the matrix $A^TA + \lambda(2\gamma_m - \alpha)I$ is positive definite, and completes the proof of (ii). The majorizer function is strictly convex when the inequality is strict. ∎

4 The MM Algorithm and Stationary Points

In this section, we prove the existence of a sequence of iteration points that are obtained by minimizing surregator functions at each iteration step. Under strongly convexity conditions for the surregators we show that the iteration points have a convergent subsequence and the limit point is a stationary point of F. Consequently, by the descending property (2.1), this stationary point will be a local minimum for F.

To prove the existence of a sequence with convergent point a local minimum, we proceed as follows. First, we need to introduce the notation for a ball. For $\epsilon > 0$ and $u \in \mathbb{R}^N$, we denote by $B_\epsilon(u)$ the ball of radius ϵ with respect to the ℓ_2 norm with center u. That is, the set of all points $x \in \mathbb{R}^N$ with ℓ_2 norm distance from the center u less than ϵ.

Proposition 4.1 *Let $\alpha > 0$ and $\epsilon > 0$. Then the sequence obtained by the following iterative algorithm converges.*

Set γ_m such that the convexity condition (3.2) holds;
Initialize $x^{(0)} \in \mathbb{R}^N$;
for $k = 0, \ldots,$ **do**
  $x^{(k+1)} = \arg\min_{x \in B_{\frac{\epsilon}{2^k}}(x^{(k)})} F^M(x, x^{(k)})$;
end
where k is the iteration counter.

Proof To prove the proposition, we first claim that the sequence $\{x^{(k)}\}_k$ has a convergent subsequence, *e.g.*, $\{x^{(k_n)}\}_n$. Then we show that the subsequence is $\{x^{(k)}\}_k$.

Boundedness: The iteration points $x^{(k)}$ satisfy

$$\|x^{(k+1)} - x^{(k)}\|_2 \le \frac{\varepsilon}{2^k}, \quad \forall k \ge 0.$$

This immediately implies that the sequence is bounded. Therefore, by The Bolzano–Weierstrass Theorem, the $\{x^{(k)}\}_k$ has a convergent subsequence with accumulation point x^*. In what follows, we prove that the sequence $\{x^{(k)}\}_k$ converges to x^*.

Convergence: Assume $\{x^{(k_n)}\}_n$ is a subsequence of $\{x^{(k)}\}_k$ such that $x^{(k_n)} \to x^*$ as $k_n \to \infty$. Fix k and let $k_n > k$. An easy calculation shows that

$$\|x^{(k)} - x^*\| \le \|x^{(k_n)} - x^*\| + \mathcal{O}\Big(\frac{\epsilon}{2^{k_n}}\Big) \quad \text{as } k \longrightarrow \infty.$$

This implies that x^* is the accumulation point for $\{x^{(k)}\}_k$ and we are done. ∎

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

Notice that the limit point cannot be a stationary or a local minimum point. However, this can be obtained under some sufficient assumptions on the majorizers. First we have a lemma.

Lemma 4.2 *Let $a > 0$ and $w \in \mathbb{R}^N$. The local majorizer $F^M(\cdot, w)$ is a-strongly convex provided that $A^T A + \lambda(\gamma_m - \alpha)I \geq 2aI$.*

Proof Recall the discrepancy of data given in (3.3):

$$F^M(x,w) = x^T\Big(\frac{1}{2}A^T A + \lambda(\gamma_m - \frac{\alpha}{2})I\Big)x + \lambda\|x\|_1 + \max_{v\in\mathbb{R}^N} g(v,x,w).$$

This representation implies that F_k^M is a-strongly convex when

$$\frac{1}{2}A^T A + \lambda\Big(\gamma_m - \frac{\alpha}{2}\Big)I \geq aI, \tag{4.1}$$

and we are done. ∎

Strong convexity is one of the most important tools in optimization and in particular it guarantees linear convergence rate of many gradient descent based algorithms. Here, we recall a result.

Lemma 4.3 ([18, Lemma B.5]) *Let f be an a-strongly convex on a convex domain $\mathcal{D}$. Let x^* be the minimizer of f on $\mathcal{D}$. Then*

$$a\|x - x^*\|_2^2 \leq f(x) - f(x^*) \quad \forall x \in \mathcal{D}.$$

As an outcome of the lemma, we prove that the limit point x^* in Theorem 4.1 is a stationary point for F, thus a local minimizer by the descending property.

Theorem 4.4 *Assume the that a-strong convexity condition (4.1) holds, and $\{x^{(k)}\}$ converges to x^*. Then x^* is a stationary point for F, and we have $\nabla F(x^*; d) \geq 0$.*

Proof By the assumption, for all $x \in \mathbb{R}^N$, we have $f_\alpha^m(x, x^{(k)}) \to f_\alpha(x) - \gamma_m\|x - x^*\|_2^2$ as $k \to \infty$. Thus, $F_k^M(x) \to F(x) + \lambda\gamma_m\|x - x^*\|_2^2$, $k \to \infty$. From the other side, by applying Lemma 4.3 to F_k^M and using its majorization property, we have

$$\begin{aligned} a\|x - x^{(k+1)}\|_2^2 &\leq F^M(x, x^{(k)}) - F^M(x^{(k+1)}, x^{(k)}) \\ &\leq F^M(x, x^{(k)}) - F(x^{(k+1)}) \quad \forall x \in \mathbb{R}^N. \end{aligned}$$

So

$$a\|x - x^{(k+1)}\|_2^2 \leq F^M(x, x^{(k)}) - F(x^{(k+1)}) \quad \forall x \in \mathbb{R}^N.$$

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

By the continuity of F, letting $k \to \infty$ in the preceding inequality, we obtain

$$a\|x - x^*\|_2^2 \le F(x) + \lambda\gamma_m\|x - x^*\|_2^2 - F(x^*),$$

or equivalently,

$$F(x) - F(x^*) \ge (a - \lambda\gamma_m)\|x - x^*\|_2^2. \tag{4.2}$$

Let $d \in \mathbb{R}^N$ be a direction with $\|d\|_2 \le \epsilon$ and $\theta > 0$. By (4.2),

$$F(x^* + \theta d) - F(x^*) \ge (a - \lambda\gamma_m)\theta^2\|d\|_2^2.$$

This implies that

$$\nabla F(x^*; d) = \liminf_{\theta \to 0^+} \frac{F(x^* + \theta d) - F(x^*)}{\theta} \ge (a - \lambda\gamma_m)\|d\|_2^2 (\liminf_{\theta \to 0^+} \theta) = 0,$$

and we are done. ■

The following result is a summary of the results that we presented in this and the previous sections.

Corollary 4.5 (Convergence) *Assume that the local majorizers $\{F^M(\cdot, x^{(k)})\}_k$ of F are a-strongly convex. The sequence of iteration points $\{x^{(k)}\}$ converges and the limit point is a local minimizer of F.*

Proof By Theorem 4.4, $\nabla F(x^*; d) \ge 0$; thus, x^* is an stationary point. By the descending property (2.1), the stationary point is a local minimum. ■

We conclude this section by illustrating some examples. First, we give some notation. For a given matrix A, we denote by $\Sigma(A)$ the set of all singular values of matrix A.

Example 4.6 (Tight frame) Assume that the rows of matrix A form a tight frame with frame constant C. Then $A^T A = CI$ and $\Sigma(A) := \{C\}$. (When $C = 1$, the rows of matrix A form a normalized tight frame, also known as Parseval frame.) Let α and λ be such that $\alpha > C\lambda^{-1}$. Then the sufficient convexity condition (2.3) fails for F, and the function F can have no local (thus global) minimum.

In the following example we present a positive lower bound for γ_m for which the convexity condition (3.2) holds for the majorizers.

Example 4.7 Assume that the convexity condition (2.3) fails. Thus, for some $\sigma \in \Sigma(A)$ we must have $\alpha > \frac{\sigma}{\lambda}$. This implies that for the smallest singular value σ_0, we also have $\alpha > \frac{\sigma_0}{\lambda}$. Define $c := \frac{\lambda\alpha - \sigma_0}{2\lambda}$. The constant c is positive, and with a straightforward computation, one can show that all pairs (γ_m, a) satisfy

$$\gamma_m \ge \frac{a}{\lambda} + c.$$

The a-strong convexity condition holds for the surregators F^M. The convexity is strict when the inequality is strict.

https://doi.org/10.4153/S0008439519000730 Published online by Cambridge University Press

References

[1] R. G. Baraniuk, E. Candes, M. Elad, and Y. Ma, eds., *Special issue on applications of sparse representation and compressive sensing*. Proc. IEEE 98(June 2010), 6.

[2] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. In: *CMS books in mathematics/Ouvrages de Mathématiques de la SMC*. Springer, New York, 2011. https://doi.org/10.1007/978-1-4419-9467-7

[3] S. O'Brien, A. N. Sinclair, and S. M. Kramer, *Recovery of a sparse spike time series by L^1 norm deconvolution*. IEEE Trans. Signal Process. 42(December 1994), 12, 3353–3365.

[4] A. Bruckstein, D. Donoho, and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*. SIAM Rev. 51(2009), 34–81. https://doi.org/10.1137/060657704

[5] E. J. Candes, M. B. Wakin, and S. Boyd, *Enhancing sparsity by reweighted $\ell 1$ minimization*. J. Fourier Anal. Appl. 14(2008), 877–905. https://doi.org/10.1007/s00041-008-9045-x

[6] R. Chartrand, E.Y. Sidky, and X. Pan, *Nonconvex compressive sensing for X-ray CT: an algorithm comparison*. In: *2013 Asilomar conference on signals, systems and computers*. IEEE, 2013, pp. 665–669.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*. SIAM J. Sci. Comput. 20(1999), 33–61. https://doi.org/10.1137/S1064827596304010

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

[9] I. Daubechies, R. DeVore, M. Fornasier, and C. Gunturk, *Iteratively reweighted least squares minimization for sparse recovery*. Comm. Pure Appl. Math. 63(2010), 1–38. https://doi.org/10.1002/cpa.20303

[10] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, *Majorization-minimization algorithms for wavelet-based image restoration*. IEEE Trans. Image Process. 16(2007), 2980–2991. https://doi.org/10.1109/TIP.2007.909318

[11] G. Gasso, A. Rakotomamonjy, and S. Canu., *Recovering sparse signals with a certain family of nonconvex penalties and DC programming*. IEEE Trans. Signal Process. 57(2009), 4686–4698. https://doi.org/10.1109/TSP.2009.2026004

[12] A. Gholami and S. M. Hosseini, *A general framework for sparsity-based denoising and inversion*. IEEE Trans. Signal Process. 59(2011), 11, 5202–5211. https://doi.org/10.1109/TSP.2011.2164074

[13] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning*. In: *Data mining, inference, and prediction*, 2nd ed., Springer Series in Statistics, Springer, New York, 2009. https://doi.org/10.1007/978-0-387-84858-7

[14] K. Lange, E. C. Chi, and H. Zhou, *A brief survey of modern optimization for statisticians*. Int. Stat. Rev. 82(2014), 46–70. https://doi.org/10.1111/insr.12022

[15] K. Lang, *Optimization*, 2nd ed., Springer Texts in Statistics, 95, Springer, New York, 2013. https://doi.org/10.1007/978-1-4614-5838-8

[16] A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari, *Nonconvex nonsmooth optimization via convex-nonconvex majorization-minimization*. Numer. Math. 136(2017), 343–381. https://doi.org/10.1007/s00211-016-0842-x

[17] D. Malioutov and A. Aravkin, *Iterative log thresholding*. In: *2014 IEEE International conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 7198–7202.

[18] J. Mairal, *Optimization with first-order surrogate functions*. In: *ICML 2013-International conference on machine learning, June 2013, JMLR proceedings, Atlanta, USA*. ICML, 2013, pp. 783–791.

[19] I. Selesnick, *Sparse regularization via convex analysis*. IEEE Trans. Signal Process. 65(2017), 4481–4494. https://doi.org/10.1109/TSP.2017.2711501

[20] I. Selesnick, *Total variation denoising via the Moreau envelope*. IEEE Signal Processing Letters 24(2017), 216–220.

[21] J. L. Starck, J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse image and signal processing: Wavelets and related geometric multiscale analysis*. Cambridge University Press, New York, 2015. https://doi.org/10.1017/CBO9781316104514

[22] H. L. Taylor, S. C. Banks, and J. F. McCoy, *Deconvolution with the l1 norm*. Geophysics 44(1979), 1, 39–52.

[23] R. Tibshirani, *Regression shrinkage and selection via the lasso*. Technical report, University of Toronto, 1994.

[24] Z. Xu, X. Chang, F. Xu, and H. Zhang., *$L_{1/2}$ regularization: A thresholding representation theory and a fast solver*. IEEE T. Neur. Net. Learn. 23(2012), 1013–1027.

[25] P. Yin, Y. Lou, Q. He, and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*. SIAM J. Sci. Comput. 37(2015), A536–A563. https://doi.org/10.1137/140952363
[26] J. Zeng, S. Lin, Y. Wang, and Z. Xu, $L_{1/2}$ *regularization: Convergence of iterative half thresholding algorithm*. IEEE Trans. Signal Process. 62(2014), 2317–2329. https://doi.org/10.1109/TSP.2014.2309076

Department of Mathematics, The City University of New York, USA
e-mail: amayeli@gc.cuny.edu