



POLITECNICO DI TORINO
Repository ISTITUZIONALE

A biconvex analysis for Lasso ℓ_1 reweighting

Original

A biconvex analysis for Lasso ℓ_1 reweighting / Fosson, Sophie Marie. - In: IEEE SIGNAL PROCESSING LETTERS. - ISSN 1070-9908. - 25:12(2018), pp. 1795-1799.

Availability:

This version is available at: 11583/2716522 since: 2018-11-07T10:56:25Z

Publisher:

Institute of Electrical and Electronics Engineers Inc.

Published

DOI:10.1109/LSP.2018.2875251

Terms of use:

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

ieee

copyright 20xx IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating .

(Article begins on next page)

A biconvex analysis for Lasso ℓ_1 reweighting

Sophie M. Fosson, *Member, IEEE*

Abstract—Iterative ℓ_1 reweighting algorithms are very popular in sparse signal recovery and compressed sensing, since in the practice they have been observed to outperform classical ℓ_1 methods. Nevertheless, the theoretical analysis of their convergence is a critical point, and generally is limited to the convergence of the functional to a local minimum or to subsequence convergence. In this letter, we propose a new convergence analysis of a Lasso ℓ_1 reweighting method, based on the observation that the algorithm is an alternated convex search for a biconvex problem. Based on that, we are able to prove the numerical convergence of the sequence of the iterates generated by the algorithm. Furthermore, we propose an alternative iterative soft thresholding procedure, which is faster than the main algorithm.

I. INTRODUCTION

In the recent literature on compressed sensing (CS, [1]), much attention has been devoted to iterative ℓ_1 reweighting (IRL1) techniques, see, *e.g.*, [2], [3], [4], [5], [6], [7], [8], [9]. In the practice, IRL1 algorithms are efficient and accurate; in theory, their convergence is guaranteed in the sense of local minimization of a certain non-convex functional, while the iterates are not proved to converge. This work goes one step further by proving the numerical convergence of the iterates, under mild conditions, leveraging a biconvex interpretation of the problem.

Hereinafter, we review the basics of CS in relation to ℓ_1 minimization (Section I-A), we introduce the IRL1 basics (Section I-B), and we illustrate the related literature along with the problem of convergence (Section I-C). Based on that, our contribution will be further specified in Section I-D.

A. Compressed sensing and ℓ_1 minimization

CS states that, under certain conditions, a k -sparse signal $x \in \mathbb{R}^n$, *i.e.*, a signal with only $k \ll n$ non-zero components, can be recovered from compressed, linear measurements $y = Ax$ (possibly corrupted by noise), where $A \in \mathbb{R}^{m,n}$, $m < n$. It is well known that CS can be re-formulated in a convex framework, using the ℓ_1 -norm to approximate ℓ_0 -norm, *i.e.*, the number of non-zeros in a vector. In particular, the Basis Pursuit (BP) and the Basis Pursuit Denoising (BPDN) formulations have been introduced, which respectively read as follows: $\min_{x \in \mathbb{R}^n} \|x\|_1$ s.t. $Ax = y$ and $\min_{x \in \mathbb{R}^n} \|x\|_1$ s.t. $\|y - Ax\|_2 < \eta$, where $\eta > 0$ is a parameter that takes account of measurement noise. (see, *e.g.*, [10, Chapter 4] for a thorough review). BP and BPDN can be solved using classic convex programming algorithms, *e.g.*, the interior points method. A non-constrained version of BPDN, known as Lasso [11], is often considered: $\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1$, with $\lambda > 0$.

Lasso was popular long before CS as variable selection method: the ℓ_1 regularization, in fact, even in overdetermined problems, can be used to find out the most relevant components in a vector. In CS, the Lasso formulation is desirable for different motivations, *e.g.*, the possibility of solution through simple iterative methods, such as the iterative soft thresholding (IST, [12], [13]) and alternating direction method of multipliers (ADMM, [14]). IST and ADMM are easy to implement and are prone to parallel computation [15] and distributed recovery, *e.g.*, in sensor networks [16], [17], [18], [19], [20].

B. Enhancing ℓ_1 minimization via reweighting

In the seminal work [2], the question arose about how ℓ_1 -minimization could be improved, starting from the observation that ℓ_1 is less "democratic" than ℓ_0 , that is, it penalizes more the coefficients with larger magnitude. This originated the attempt to reweight the ℓ_1 -norm term in order to improve the final estimation. The proposal in [2] was to use weights that are inversely proportional to the magnitudes of the corresponding coefficients. This was implemented through the following IRL1 algorithm for BP: given $x(0) \in \mathbb{R}^n$, for $t = 1, \dots, T_{stop}$,

$$w_i(t) = \frac{1}{|x_i(t)| + \varepsilon} \text{ for any } i \in \{1, \dots, n\} \quad (1)$$

$$x(t+1) = \operatorname{argmin}_{x \in \mathbb{R}^n} \sum_{i=1}^n w_i(t) |x_i| \text{ s.t. } y = Ax$$

where $\varepsilon > 0$ is necessary to avoid infinite weights, and T_{stop} is the time step at which a suitable convergence criterion is satisfied. This algorithm converges in the sense that given $g(|x_i|) := \log(|x_i| + \varepsilon)$, the functional $\sum_{i=1}^n g(|x_i(t)|)$ converges to a (local) minimum of

$$\sum_{i=1}^n g(|x_i|) \text{ s.t. } Ax = y. \quad (2)$$

This was proved exploiting a Minimization-Maximization (MM, [21]) perspective, as explained in [2, Section 2.3]. Nevertheless, the iterates $x(t)$ are not guaranteed to converge. In the practice, this is a drawback, because it requires the algorithm to compute $\sum_{i=1}^n g(|x_i(t)|)$ to evaluate T_{stop} . If the iterates $x(t)$ were known to converge, one should simply check $|x_i(t) - x_i(t-1)| < \delta$ for a given tolerance δ , singularly for each i , without computing the whole functional.

C. Review of the literature and the convergence problem

Beyond [2], reviewed in the previous section, other works have focused on iterative reweighting in the last decade.

It is worth to mention that a popular class of reweighting methods is the so-called iteratively reweighted least squares (IRLS, [22]). IRLS techniques perform an ℓ_2 reweighting,

i.e., they adaptively modify the weights of a least squares functional, which has been proved to be efficient for sparse recovery [23], [24], [25]. Their convergence is a tricky point: in general, they converge only to a set of points [26], while stronger convergence has been proved only in specific settings.

Coming back to ℓ_1 reweighting, in [27] the approach of [2] was analyzed in the noisy case. In [3], algorithm (1) was re-elaborated using weights $(|x_i(t)| + \epsilon(t))^{q-1}$, with $q \in (0, 1)$ and non-increasing $\epsilon(t)$. [3, Proposition 4.2] proves the existence of a convergent subsequence. Convergence to the correct sparse signal was shown in numerical experiments. The same method was analyzed in [8], which provided some further characterization of the accumulation points.

In [5], a unified theoretical analysis was proposed for a large class of IRL1 methods for noise-free, underdetermined linear systems. Convergence analyses were proposed for different concave *merit* functions $F_\epsilon(x) = \sum_i g_i(|x_i| + \epsilon)$ [5, Eq 2.1], under some technical assumptions [5, Assumption 2.1]. In general, convergence of the functional to a local minimum and subsequence convergence were achieved. For $F_\epsilon(x)$ bounded below for any $|x_i| > 0$ and $\epsilon > 0$, numerical convergence was also proved for an IRL1 procedure with decreasing ϵ [28, Algorithm 2.2, Corollary 3.4]. However, this is not valid in classical cases, *e.g.*, for algorithm (1). Moreover, vanishing parameters are generally not desired in all those applications where the signal is not finite in time and processing/recovery must be performed continuously.

In [7], homotopy strategies were elaborated for the problem $\min \sum_{i=1}^n w_i |x_i| + \frac{1}{2} \|Ax - y\|_2^2$, whose efficiency was verified via numerical experiments. Finally, in [9], composite regularization was considered, and convergence was proved in the sense of asymptotic stationary point condition [29, Definition 2.4], leveraging an MM interpretation.

In conclusion, to the best of our knowledge, previous convergence analyses for IRL1 algorithms generally achieved the convergence of the functional to a stationary point or local minimum, while numerical convergence was shown only in a specific case with decreasing parameter ϵ [5]. We refer the interested reader to [4], [6] for further comparisons between ℓ_1 (IRL1) and ℓ_2 (IRLS) reweighting methods.

We finally notice that problem (2) is concave. In the following, we will see that ℓ_1 reweighting can tackle sparse optimization problems with concave penalization. The interest on sparse optimization with concave penalization has substantially increased in the last few years, and novel non-reweighting techniques have been proposed [30], [31], [20], [33]. However, most of this literature [30], [31] does not envisage the compressed case (as it requires a non-singular $A^T A$). For reasons of space, we leave for future work an extended analysis of non-compressed case along with the comparison to novel non-reweighting approaches.

D. Contribution

In this letter, we propose a biconvex interpretation of a Lasso ℓ_1 reweighting method. This new perspective allows us to prove the numerical convergence of the iterates, that is, $\lim_{t \rightarrow \infty} \|x(t) - x(t-1)\|_2 = 0$. This is not sequence

convergence, *i.e.*, we are not proving that $x(t) \rightarrow x^* \in \mathbb{R}^n$. However, numerical convergence is sufficient for practical purposes, for example to evaluate T_{stop} .

The remainder of the letter is organized as follows. In Section II, we state the main problem, the main algorithm, the biconvex interpretation, and our convergence results. Afterwards, in Section III, we will propose a simplified algorithm which overcomes the problem of nested loops in the main algorithm and requires less iterations.

II. PROBLEM STATEMENT AND THEORETICAL ANALYSIS

As mentioned in [2], problem (2) can be formulated for any concave, non-decreasing $g(|x_i|)$ [34]. In addition to $\log |x_i| + \epsilon$, other popular g 's are ℓ_q , with $q \in (0, 1)$ [24], [22], [25], [34]; smoothly clipped absolute deviations (SCAD) [35]; minimax concave penalty (MCP) [36], [34], [20], [37]. Extending [2] to generic concave, non-decreasing g 's, the following algorithm can be implemented to achieve a (local) minimum of $\sum_{i=1}^n g(|x_i|)$ s.t. $y = Ax$: given $x(0) \in \mathbb{R}^n$, for any $t = 1, \dots, T_{stop}$

$$\begin{aligned} w_i(t) &= g'(|x_i(t)|) \text{ for any } i \in \{1, \dots, n\} \\ x(t+1) &= \operatorname{argmin}_{x \in \mathbb{R}^n} \sum_{i=1}^n w_i(t) |x_i| \quad \text{s.t. } y = Ax \end{aligned} \quad (3)$$

The analytical justification provided in [2] is as follows. Since $G = \sum_{i=1}^n g(|x_i|)$ is concave, thus below its tangent, one can improve on a guess $v \in \mathbb{R}^n$ by locally minimizing a linearization of G around v . In this way, one obtains an MM procedure which corresponds to the algorithm (3). Leveraging this interpretation, the convergence of the algorithm can be proved using the Global Convergence Theorem [38, Chapter 7], as also illustrated in [39]. As remarked in [2, Section 2.3], convergence is here intended to a local minimum of the functional, while the iterates of the algorithm are not proved to converge. Following this philosophy, let us reformulate the Lasso problem using a concave penalty (from now onwards, $\mathbb{R}_+ = [0, +\infty)$):

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \sum_{i=1}^n g(|x_i|), \quad \lambda > 0 \quad (4)$$

$g: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ concave, non-decreasing in $|x_i|$.

To tackle problem (4), we propose Algorithm 1, which is an analogous for Lasso of the algorithm (3) for BP. The conver-

Algorithm 1 Lasso IRL1

- 1: Initialize: $x(0) \in \mathbb{R}^n$; $\lambda > 0$
 - 2: **for all** $t = 1, \dots, T_{stop}$ **do**
 - 3: $w_i(t) = g'(|x_i(t)|)$ for any $i \in \{1, \dots, n\}$
 - 4: $x(t+1) = \operatorname{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \sum_{i=1}^n w_i(t) |x_i|$
 - 5: **end for**
-

gence of Algorithm 1 cannot be derived with the techniques used for algorithm 3, as the functional (4) is the sum of a convex and a concave terms, thus in general it is neither convex nor concave. The minimization on the local linearization then

does not guarantee the descent. We now provide a convergence proof for Algorithm 1, based on biconvex minimization tools.

A. Biconvex perspective

Following the notation in [40], we remind that a function $f : X \times Y \rightarrow \mathbb{R}$ is *biconvex* if it is convex in y (respectively, in x) for any fixed $x \in X$ (respectively, $y \in Y$). The problem

$$\min f(x, y) \quad (x, y) \in B, \quad B \subseteq X \times Y, \quad B \text{ biconvex}$$

is called a *biconvex optimization problem* if f is biconvex on B . We say that $(x^*, y^*) \in B$ is a *partial optimum* of f on B if $f(x^*, y^*) \leq f(x, y^*)$, for any $x \in B_{y^*}$, and $f(x^*, y^*) \leq f(x^*, y)$, for any $y \in B_{x^*}$ where B_x and B_y respectively are the x -section and the y -section of B .

We name alternated convex search (ACS, [40, Algorithm 4.1]) the algorithm that iteratively minimizes $f(x, y)$ with respect to x and y . Each iteration corresponds then to the solution of a convex problem, namely, $x(t+1) = \min_{x \in B_{y(t)}} f(x, y(t))$; $y(t+1) = \min_{y \in B_{x(t+1)}} f(x(t+1), y)$.

We can now prove our main results.

Theorem 1. *Let $g(|x_i|)$ be strictly concave, increasing, and differentiable in its domain $D \subseteq \mathbb{R}_+$ (for $|x_i| = 0$, the right derivative is considered). Let us define the functional:*

$$\begin{aligned} \mathcal{F} : X \times Y &\rightarrow \mathbb{R}, \quad X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}_+^n \\ \mathcal{F}(x, w) &:= \frac{1}{2} \|y - Ax\|_2^2 + \lambda \sum_{i=1}^n [w_i |x_i| + h(w_i)] \end{aligned} \quad (5)$$

where $h : H \subseteq \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is defined by $h' = -(g')^{-1}$, $(\cdot)^{-1}$ indicating the inverse function. Then, \mathcal{F} is biconvex and Algorithm 1 is an ACS for it.

Before proving the theorem, let us discuss the definitions of h , X , Y for some popular g 's.

a) If $g(|x_i|) = \log(|x_i| + \epsilon)$ [2], [32], $g'(|x_i|) = (|x_i| + \epsilon)^{-1}$. Computing its inverse, we easily get $h'(w_i) = \epsilon - \frac{1}{w_i}$, from which $h(w_i) = \epsilon w_i - \log(w_i)$, modulus an integration constant which can be arbitrarily set (for instance, one can choose it so that $h(w_i) \geq 0$ to have a non-negative \mathcal{F}). Since the corresponding \mathcal{F} is bounded below, continuous and coercive, given any compact set $S \subset \mathbb{R}^n \times \mathbb{R}^n$ of starting points, Algorithm 1 evolves in the compact set $\{(x, w) : \mathcal{F}(x, w) \leq \max_{s \in S} \mathcal{F}(s)\}$. We can then assume that $X = [-\beta, \beta]^n$ for some sufficiently large $\beta > 0$; thus, $Y = [(\beta + \epsilon)^{-1}, \epsilon^{-1}]^n$.

b) If $g(|x_i|) = (|x_i| + \epsilon)^q$, $q \in (0, 1)$ [3], we obtain $h(w_i) = \epsilon w_i + \frac{1-q}{(w_i/q)^{q/(1-q)}}$. Again, we can set $X = [-\beta, \beta]^n$, $Y = [q/(\beta + \epsilon)^{1-q}, q/\epsilon^{1-q}]^n$.

c) If $g(|x_i|) = \alpha |x_i| - \frac{1}{2} |x_i|^2$ with $\alpha > 0$ and $X = [-\alpha, \alpha]^n$ [36], [20], we can obtain $h(w_i) = \frac{1}{2}(\alpha - w_i)^2$, $Y = [0, \alpha]^n$.

Proof. First of all, since g is strictly concave, the derivative g' is strictly monotonically decreasing. Therefore g' is invertible and h' is well defined. Moreover, h' turns out to be strictly monotonically increasing, thus h is strictly convex. As a consequence, \mathcal{F} is convex in w . As the convexity in x is evident, we conclude that \mathcal{F} is biconvex.

Regarding the ACS of \mathcal{F} , we notice that the minimization with respect to x simply is a (weighted) Lasso, as in the

first step of Algorithm 1. To minimize with respect to w , we derive with respect to w_i , $i \in \{1, \dots, n\}$, and observe that the gradient is equal to zero whenever $|x_i| = -h'(w_i)$ for any i . By inversion we obtain $w_i = g'(|x_i|)$, which corresponds to the second step of Algorithm 1. This proves the thesis. \square

We specify that (5) is not derived from (4), and the two functionals are not said to be equivalent. The properties of (5) as cost functional are not of our interest: we instead exploit (5) to study the convergence properties of Algorithm 1.

The fact that Algorithm 1 is an ACS for a biconvex functional yields to the following theorem.

Theorem 2. *Let \mathcal{F} be as defined in (5). Let us assume that X and Y are closed, and that Lasso in Algorithm 1 has a unique solution. Then, Algorithm 1 achieves a partial optimum of \mathcal{F} . Moreover, the iterates of Algorithm 1 numerically converge, that is, $\lim_{t \rightarrow \infty} \|(x(t+1), w(t+1)) - (x(t), w(t))\|_2 = 0$.*

Theorem 2 is a consequence of the following result.

Theorem 3. *([40, Theorem 4.9]) Let $f : X \times Y \rightarrow \mathbb{R}$ be biconvex, continuous, and let X and Y be closed. If the sequence generated by ACS is contained in a compact set, then it has a least one accumulation point (x^*, y^*) . Moreover, if the convex problems $\min_x f(x, y^*)$ and $\min_y f(x^*, y)$ both admit unique solutions, then all the accumulation points are partial optima and they form a connected compact set, and the sequence is numerically convergent.*

Concerning the closedness of X and Y , we have previously discussed in the examples a)-b)-c) the fact that for popular g 's we can work on compact sets. Moreover, uniqueness for the Lasso step is guaranteed for most used sensing matrices, e.g., random sensing matrices generated from continuous distributions (see [41]). On the other hand, $h(w_i)$ is strictly convex, hence \mathcal{F} has a unique minimum with respect to each w_i .

In conclusion, Theorem 2 is valid for a wide class of popular IRL1 methods.

III. A SIMPLE IST VARIANT

The first step of Algorithm 1 requires to solve a Lasso problem, which can be performed via iterative methods like IST and ADMM. This however entails a nested loop structure, which is generally not desired. We then propose a variant, summarized in Algorithm 2, in which the solution of Lasso is substituted by a single IST step. $\mathbb{S}_{\lambda w}$ indicates the soft thresholding operator [13] with componentwise threshold λw_i .

Algorithm 2 Lasso IRL1 IST

- 1: Initialize: $x(0) \in \mathbb{R}^n$, $\lambda > 0$ $\tau > 0$
 - 2: **for all** $t = 1, \dots, T_{stop}$ **do**
 - 3: $w_i(t) = g'(|x_i(t)|)$ for any $i \in \{1, \dots, n\}$
 - 4: $x(t+1) = \mathbb{S}_{\lambda w(t)}[x(t) + \tau A^T(y - Ax(t))]$
 - 5: **end for**
-

Proposition 1. *Assume that $\tau \|A\|_2^2 < 1$ in Algorithm 2. Given a sequence $(x(t), w(t))$ generated by Algorithm 2, then $\mathcal{F}(x(t), w(t))$ converges to a partial optimum.*

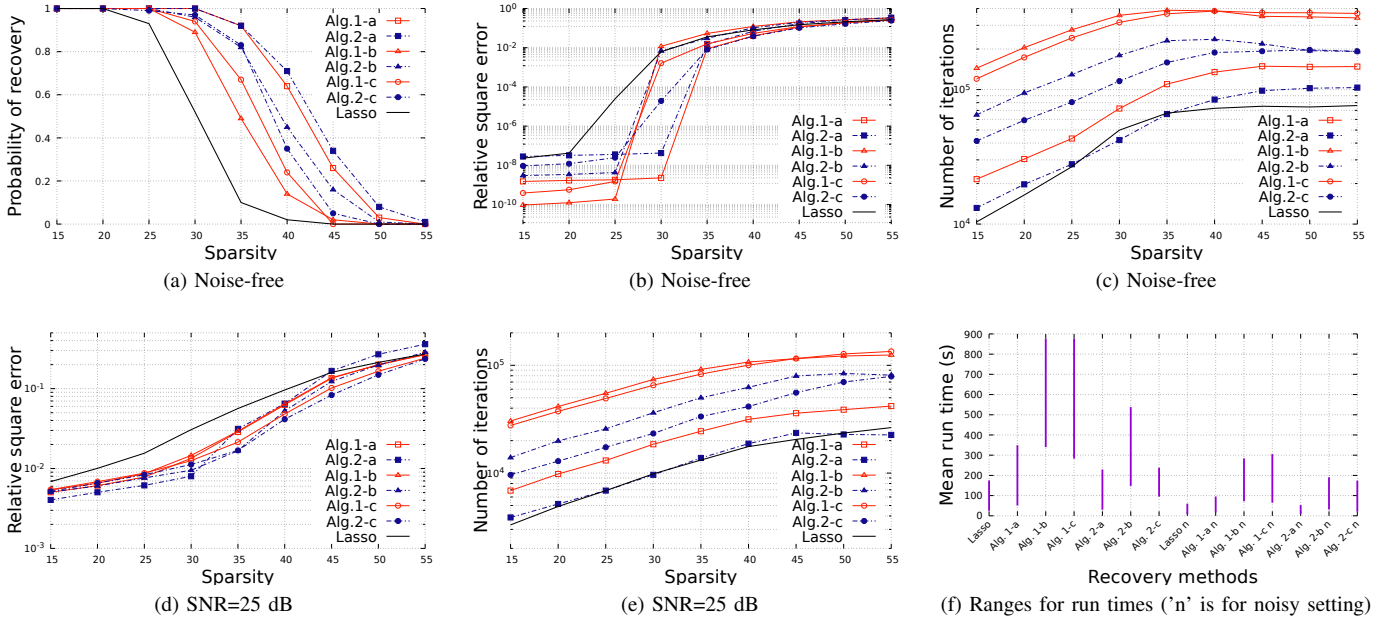


Figure 1: Lasso (via ADMM) vs Algorithms 1 and 2; $n = 256$, $m = 100$.

Proof. Algorithm 2 corresponds to the alternating minimization of the tri-convex surrogate functional $\mathcal{H}(x, w, b) = \mathcal{F}(x, w) + \frac{1}{2} \|x - b\|_2^2 - \frac{\tau}{2} \|A(x - b)\|_2^2$ adding the step $b(t) = x(t)$ (see [13], for a review on the surrogate functional idea). For any block sub-problem, the minimum is unique under conditions discussed above, therefore $\mathcal{H}(x(t), w(t), b(t))$, and thus $\mathcal{F}(x(t), w(t))$, is strictly decreasing outside the solution set. This implies the thesis. \square

In [42], [43], more insight on alternating minimization/block coordinate descent of multi-convex or non-convex, non-differentiable problems can be found.

A. Numerical experiment

We finally show a numerical experiment to practically verify the efficiency of our algorithms.¹ We consider the experiment proposed in [2, Section 3.1], whose setting is as follows: $n = 256$, $m = 100$, $k \in [15, 55]$; the unknown sparse signal has support generated uniformly at random, while its non-zero positions are randomly generated from a standard Gaussian distribution $\mathcal{N}(0, 1)$; the entries of the sensing matrix A are randomly generated from a Gaussian distribution $\mathcal{N}(0, \frac{1}{m})$. We consider the penalties g 's defined in a), b), and c) in Section II-A with $\varepsilon = 10^{-1}$, $q = 1/2$, $\alpha = 2$.

We use ADMM to solve the classical Lasso and the Step 3 of Algorithm 1. We set $x(0) = (0, \dots, 0)^T$, $\tau = 2.5 \times 10^{-1}$. We consider the noise-free case and the case when the signal-to-noise ratio is SNR=25dB; λ is respectively set to 10^{-5} and 10^{-4} . The algorithms are stopped whenever $\|x(t) - x(t-1)\|_2 < 10^{-5}$. At most two reweighting iterations are performed for Algorithm 1. As noticed in [2], in fact, the most of the enhancement is obtained within the first few reweighting iterations. The shown results are averaged over 100 runs.

¹The code to reproduce this experiment is available on <https://github.com/sophie27/Lasso-11-reweighting>.

In Figure 1 we show the performance of the proposed algorithms compared to Lasso, varying the sparsity level. Let \tilde{x} and \hat{x} be the original signal and the estimation, respectively. The performance metrics are: the empirical probability of recovery for the noise-free case (as in [2], we define recovery the case $\|\tilde{x} - \hat{x}\|_\infty < 10^{-3}$); the relative square error $\frac{\|\tilde{x} - \hat{x}\|_2^2}{\|\tilde{x}\|_2^2}$; the total number of iterations and run time. Simulations are performed on an AMD Opteron 6276, 2.30 GHz Processor.

In Figure 1, we can appreciate a general improvement obtained by Algorithms 1 and 2 with respect to Lasso in terms of recovery and relative square error, in particular using the log penalty a). In the noise-free case, the smaller λ generates a more precise solution, but yields to slower convergence. In the noisy case, the larger λ helps to tolerate noise and at the same time accelerates the procedure. Figure 1(d) shows the ranges (in seconds) of the mean run times for each considered k . This is consistent with the number of iterations (Figure 1(c)-(e)).

Clearly, Lasso requires less iterations than Algorithm 1, which is more accurate at the price of a longer run time. Algorithm 2 reduces the run time without substantial loss of accuracy. Algorithm 2-a is the best choice for the proposed experiment: it has the best recovery accuracy in the noise free case (see Figure 1(a)) and run time very close to Lasso; it only loses some accuracy for large k in the noisy setting.

IV. CONCLUSIONS

In this letter, we have proved that a Lasso iterative ℓ_1 reweighting algorithm corresponds to the alternating minimization of a biconvex functional. This allows us to prove the numerical convergence of the iterates of the algorithm, that is, the distance between two successive iterates goes to zero. This is a stronger convergence with respect to previous results on iterative ℓ_1 reweighting methods. Moreover, we have proposed an IST-based alternative algorithm, which converges to a partial optimum and is practically very fast.

REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journ. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [3] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 395–407, 2009.
- [4] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 317–329, April 2010.
- [5] Y.-B. Zhao and D. Li, "Reweighted ℓ_1 -minimization for sparse solutions to underdetermined linear systems," *SIAM J. Optim.*, vol. 22, no. 3, pp. 1065–1088, 2012.
- [6] Q. Lyu, Z. Lin, Y. She, and C. Zhang, "A comparison of typical ℓ_p minimization algorithms," *Neurocomput.*, vol. 119, pp. 413–424, 2013.
- [7] M. Asif and J. Romberg, "Fast and accurate algorithms for re-weighted ℓ_1 -norm minimization," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5905–5916, 2013.
- [8] X. Chen and W. Zhou, "Convergence of the reweighted ℓ_1 minimization algorithm for ℓ_2 - ℓ_p minimization," *Comput. Optim. Appl.*, vol. 59, no. 1-2, pp. 47–61, 2014.
- [9] R. Ahmad and P. Schniter, "Iteratively reweighted ℓ_1 approaches to sparse composite regularization," *IEEE Trans. Computat. Imag.*, vol. 1, no. 4, pp. 220–235, 2015.
- [10] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.
- [11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1996.
- [12] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413 – 1457, 2004.
- [13] M. Fornasier, "Numerical methods for sparse recovery," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, M. Fornasier, Ed. Radon Series Comp. Appl. Math., de Gruyter, 2010, pp. 93–200.
- [14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1 – 122, 2010.
- [15] A. Fianndrotti, S. M. Fosson, C. Ravazzi, and E. Magli, "GPU-accelerated algorithms for compressed signals recovery with application to astronomical imagery deblurring," *Int. J. Remote Sens.*, vol. 39, no. 7, pp. 2043–2065, 2018.
- [16] J. Mota, J. Xavier, P. Aguiar, and M. Puschel, "D-ADMM: a communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2718–2723, 2013.
- [17] S. M. Fosson, J. Matamoros, C. Antón-Haro, and E. Magli, "Distributed support detection of jointly sparse signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 6434–6438.
- [18] C. Ravazzi, S. M. Fosson, and E. Magli, "Distributed iterative thresholding for ℓ_0/ℓ_1 -regularized linear inverse problems," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2081–2100, 2015.
- [19] J. Matamoros, S. M. Fosson, E. Magli, and C. Antón-Haro, "Distributed ADMM for in-network reconstruction of sparse signals with innovations," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 4, pp. 225–234, 2015.
- [20] S. M. Fosson, J. Matamoros, C. Antón-Haro, and E. Magli, "Distributed recovery of jointly sparse signals under communication constraints," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3470–3482, 2016.
- [21] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2017.
- [22] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.
- [23] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, 2007.
- [24] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 3869–3872.
- [25] C. Ravazzi and E. Magli, "Gaussian mixtures based IRLS for sparse recovery with quadratic convergence," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3474–3489, 2015.
- [26] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 183–195, 2014.
- [27] D. Needell, "Noisy signal recovery via iterative reweighted ℓ_1 -minimization," in *Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 113–117.
- [28] C.-H. Zhang and T. Zhang, "A general theory of concave regularization for high-dimensional sparse estimation problems," *Statist. Sci.*, vol. 27, no. 4, pp. 576 –593, 2012.
- [29] J. Mairal, "Incremental majorization-minimization optimization with application to large-scale machine learning," *SIAM J. Optim.*, vol. 25, no. 2, pp. 829–855, 2015.
- [30] A. Parekh and I. W. Selesnick, "Convex denoising using non-convex tight frame regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1786–1790, 2015.
- [31] Á. Bayram, "On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1597–1608, 2016.
- [32] M. Calvo-Fullana, J. Matamoros, C. Antón-Haro, and S. M. Fosson, "Sparsity-promoting sensor selection with energy harvesting constraints," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3766–3770.
- [33] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, 2017.
- [34] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Problems*, vol. 32, no. 7, pp. 75 004–75 028, 2016.
- [35] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [36] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
- [37] X. Huang and M. Yan, "Nonconvex penalties with analytical solutions for one-bit compressive sensing," *Signal Process.*, vol. 144, pp. 341–351, 2018.
- [38] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 4th ed. Springer International Publishing Switzerland, 2016.
- [39] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *IEEE Proc. American Control Conference (ACC)*, vol. 3, 2003, pp. 2156–2162.
- [40] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Mathematical Methods of Operations Research*, vol. 66, no. 3, pp. 373–407, 2007.
- [41] R. J. Tibshirani, "The Lasso problem and uniqueness," *Electronic Journal of Statistics*, vol. 7, pp. 1456–1490, 2013.
- [42] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, no. 6, pp. 1758–1789, 2013.
- [43] —, "A globally convergent algorithm for nonconvex optimization based on block coordinate update," *J. Sci. Comput.*, vol. 72, no. 2, pp. 700–734, 2017.