A NEW GENERALIZED THRESHOLDING ALGORITHM FOR INVERSE PROBLEMS WITH SPARSITY CONSTRAINTS

Sergey Voronin[*]

Université de Nice Sophia-Antipolis
CNRS (UMR 6526), IRD
Obs. de la Côte d'Azur, Géoazur
250 avenue Albert Einstein 06560 Valbonne, France

Rick Chartrand[†]

Los Alamos National Laboratory
Los Alamos, NM 87545

ABSTRACT

We propose a new generalized thresholding algorithm useful for inverse problems with sparsity constraints. The algorithm uses a thresholding function with a parameter p, first mentioned in [1]. When $p = 1$, the thresholding function is equivalent to classical soft thresholding. For values of p below 1, the thresholding penalizes small coefficients over a wider range and applies less bias to the larger coefficients, much like hard thresholding but without discontinuities. The functional that the new thresholding minimizes is non-convex for $p < 1$. We state an algorithm similar to the Iterative Soft Thresholding Algorithm (ISTA) [2]. We show that the new thresholding performs better in numerical examples than soft thresholding.

Index Terms— inverse problems, sparsity, thresholding, compressive sensing, image denoising

1. INTRODUCTION

Many applications involve at some stage, a linear inverse problem $\bar{A}\bar{x} = \bar{b}$ with $\bar{A} \in \mathbb{R}^{m\times N}$ and $\bar{b} \in \mathbb{R}^m$. When $m < N$, the system is underdetermined and there are infinitely many solutions. In addition, the matrix $\bar{A}$ and/or the vector $\bar{b}$ may be unavailable, instead we are given the noisy versions A and b. We then replace the system $\bar{A}\bar{x} = \bar{b}$ by the least squares problem $\min_x \|Ax - b\|_2$. Again, when the system is underdetermined, the least squares problem has many solutions. Thus, additional constraints must be imposed to make the problem well posed. The classical constraint is on the ℓ_2 norm of the solution, resulting in the quadratic minimization problem $\min_x \|Ax-b\|_2^2+\lambda\|x\|_2^2$. Many applications, however, lend themselves to sparse solutions. Sparsity means having few nonzero coefficients with respect to a given basis. The ℓ_2 constraint does not give sparse solutions. Instead, we must use a different constraint. The measure of nonzeros, $\|\cdot\|_0$, is not a norm, highly non-convex, and difficult to deal with numerically. For these reasons, other penalty functions have been proposed. Among them, the closest convex norm to the ℓ_0 penalty is the ℓ_1 norm: $\|x\|_1 = \sum_{k=1}^N |x_k|$. In many cases, this constraint gives sparse solutions [3, 4].

This reasoning motivates the minimization of the convex but non-smooth ℓ_1 functional $\|Ax - b\|_2^2 + 2\tau\|x\|_1$. The non-smoothness makes minimizing this functional computationally challenging. For large-scale problems, methods that require a linear system to be solved at every iteration become infeasible. One approach that avoids this problem is the Iterative Soft-Thresholding Algorithm (ISTA) [2]. Each iteration of this algorithm is computationally efficient, while the iterates are guaranteed to converge.

In this paper, we consider more general penalty functions than the ℓ_1 norm, motivated by many previous results showing that nonconvex minimization gives much better performance for compressive sensing [5, 6, 7] and image restoration [8]. We propose generalizations of ISTA, using modified thresholding functions that minimize non-convex functions. In the next section, we define the thresholding functions and the penalty functions that they minimize. In Sec. 3, we describe the corresponding algorithms, and present its theoretical properties. In Sec. 4, we present numerical results showing the improved performance of the new algorithms.

2. THRESHOLDING FUNCTIONS

The well known soft thresholding operator is defined as:

$$S_\tau(x) = \begin{cases} x - \tau & \text{if } x \geq \tau \\ 0 & \text{if } -\tau \leq x \leq \tau \\ x + \tau & \text{if } x \leq -\tau \end{cases}$$

with the operation applied component-wise to each entry of a vector. The operation can be written as:

$$(S_\tau(x))_k = \operatorname{sgn}(x_k)\max\{0, |x_k| - \tau\}$$

The connection to the ℓ_1 minimization problem is that the functional $\|Ax - b\|_2^2 + 2\tau\|x\|_1$ can be minimized using the

[*]This work was supported in part by ERC Advanced Grant 226837.

[†]This work was supported by the U.S. Department of Energy through the LANL/LDRD Program.

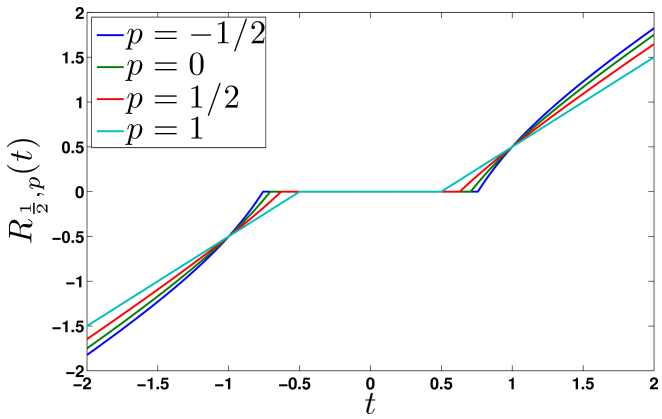

Fig. 1. Plot of p-thresholding function for a few values of p. Smaller values of p shrink more values to zero, while shrinking large values less.

simple iterative scheme ISTA:

$$x^{n+1} = S_\tau\big(x^n + A^T b - A^T A x^n\big) \tag{1}$$

for any initial guess x^0 as long as the spectral norm $\|A\|_2 < 1$. In this paper, we use the generalized thresholding operator defined via:

$$(R_{\tau,p}(x))_k = \operatorname{sgn}(x_k)\max\{0, |x_k| - \tau|x_k|^{p-1}\} \tag{2}$$

Examples are plotted in Fig. 1. We observe that the new thresholding penalizes small coefficients more as the value of p is reduced. Additionally, the new thresholding function with $p < 1$ applies less bias to the large coefficients.

We now briefly discuss how the new thresholding function arises. We recall the following property of soft thresholding:

$$S_\tau(t) = \arg\min_s (s-t)^2 + 2\tau|s|.$$

This property gives rise to the following Huber function:

Lemma 2.1. *The minimum value of* $\frac{1}{2\tau}(s-t)^2 + |s|$ *is:*

$$h_\tau(t) = \begin{cases} \frac{t^2}{2\tau} & \text{if } |t| \le \tau \\ |t| - \frac{\tau}{2} & \text{if } |t| > \tau. \end{cases}$$

Now we generalize the above to the following function:

$$\bar{h}_{\tau,p}(t) = \begin{cases} \frac{t^2}{2\tau} & \text{if } |t| \le \tau^{\frac{1}{2-p}} \\ \frac{|t|^p}{p} - \sigma & \text{if } |t| > \tau^{\frac{1}{2-p}}. \end{cases}$$

Notice that when $p = 1$, we have that $\bar{h}_{\tau,1}(b) = h_\tau(b)$. When $p = 0$, we interpret $|t|^p/p$ to mean $\log|t|$. We choose σ to make $\bar{h}$ continuous, with the result that $\bar{h} \in C^1$.

We now make use of the *Legendre-Fenchel* transform given by:

$$f^*(y) = \max_x \langle x, y\rangle - f(x). \tag{3}$$

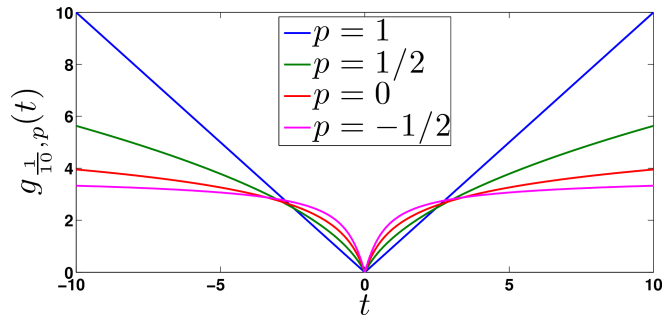

Fig. 2. Plot of $g_{\tau,p}$ for $\tau = 0.1$ and a few values of p.

We introduce the function $g_{\tau,p}$ defined by:

$$\frac{|s|^2}{2} + \tau g_{\tau,p}(s) = \left(\frac{|\cdot|^2}{2} - \tau h_{\tau,p}(\cdot)\right)^*(s) \tag{4}$$

and arrive at the following property of the new thresholding function:

Lemma 2.2. *The thresholding function* $R_{\tau,p}$ *satisfies the minimization problem:*

$$R_{\tau,p}(b) = \arg\min_x (x-b)^2 + 2\tau g_{\tau,p}(x)$$

We note that the minimizer in Lemma 2.2 is unique. Indeed, simple manipulation shows the argmin of (2.2) is the argmax of (3) with $f = \varphi^*$ where $\varphi(t) = \frac{|t|^2}{2} - \tau h_{\tau,p}(t)$ is the function used to define g in (4). By [9, Prop. 11.3], this argmax is the same set as $\partial\varphi(y)$, which is a singleton because $\varphi \in C^1$.

It is only possible to compute $g_{\tau,p}(t)$ explicitly for special values of p (like $p = 1/2$, where it amounts to solving a cubic equation). However, our algorithm will not require computing $g_{\tau,p}(t)$, it will only require finding the minimizer of expressions of the form (2.2). And this, by design, is done using the simple p-thresholding function. Fig. 2 shows numerically computed plots of $g_{\tau,p}$ for a few values of p. It grows like $|t|^p/p$ for large $|t|$, in particular being bounded above for $p < 0$. Near $t = 0$, the graphs have a sharp corner, but not a vertical tangent like $|t|^p$ does.

3. THE GENERALIZED THRESHOLDING ALGORITHM

Following the development of ISTA, we define the following algorithm using the generalized thresholding function $R_{\tau,p}$:

$$x^{n+1} = R_{\tau,p}\big(x^n + A^T b - A^T A x^n\big) \tag{5}$$

Like ISTA, the algorithm can be used from any initial point x^0 as long as $\|A\|_2 < 1$ (which can be accomplished by rescaling by the largest singular value of A). We assume this is true henceforth.

Proposition 3.1. *Any fixed point of* (5) *is a local minimizer of the functional*

$$F(x) = \|Ax - b\|_2^2 + 2\tau G_{\tau,p}(x), \tag{6}$$

where $G_{\tau,p}(x) = \sum_{k=1}^{N} g_{\tau,p}(x_k)$.

We omit the proof for lack of space, but the main fact is that although $G_{\tau,p}$ is nonconvex, it is *subdifferentially regular*, so that first-order optimality conditions are sufficient for a point to be a local minimizer (see [9, Thm. 10.1]).

Many properties of the algorithm follow from standard results of majorization-minimization (MM) algorithms [10]. Such algorithms require a function M such that for all x and y, $F(x) = M(x,x)$ and $F(x) \leq M(x,y)$. Then one sets $x^{n+1} = \arg\min_x M(x, x^n)$. One immediately obtains that $F(x^{n+1}) \leq F(x^n)$. We obtain the algorithm (5) from

$$M(x,y) = \|Ax-b\|_2^2 - \|A(x-y)\|_2^2 + \|x-y\|_2^2 + 2\tau G_{\tau,p}(x).$$

This construction allows us to prove useful properties of the algorithm:

Lemma 3.2. *The iterates* (x^n) *satisfy* $\|x^{n+1} - x^n\|_2 \to 0$. *Also, for* $p > 0$ *we have that* $\|x^n\|_2$ *is bounded.*

Proof. The first claim follows just as in the case of ISTA; see [2]. The second follows from the fact that $F(x^n)$ is nonincreasing, and that $G_{\tau,p}$ (and hence F) is coercive when $p > 0$. □

Next we show that the limit points of all converging subsequences of (x^n) are local minimizers:

Lemma 3.3. *Every convergent subsequence of* (x^n) *converges to a local minimizer of* $F(x)$.

Note that by Lemma 3.2, the set of convergent subsequences is nonempty.

Proof. By Proposition 3.1, we need only show that the limit of a convergent subsequence is a fixed point. Take any convergent subsequence $x^{n_j} \to \bar{x}$. By Lemma 3.2, we have that $x^{n_j+1} \to \bar{x}$ also. By continuity of the thresholding $R_{\tau,p}$,

$$\begin{aligned} \bar{x} &= \lim x^{n_j+1} = \lim R_{\tau,p}(x^{n_j} + A^T b - A^T A x^{n_j}) \\ &= R_{\tau,p}(\bar{x} + A^T b - A^T A \bar{x}). \end{aligned} \tag{7}$$

Thus $\bar{x}$ is a fixed point of (5). □

This leads us to our main theorem.

Theorem 3.4. *Let* $p > 0$. *Then the algorithm* (5) *is globally convergent to a local minimizer of* F.

We do not have space to present the proof. The main tool is a result of Meyer [11] that establishes that for the algorithm not to be globally convergent implies that the set of limit points forms a continuum. The proof proceeds by showing that having first-order optimality conditions holding on a continuum implies third-order properties of $g_{\tau,p}$ that are false.

We also note that while our proof of convergence only holds for $p > 0$, in practice the algorithm converges quite well when $p < 0$, with this case often producing the best numerical results.

Our algorithm follows:

Algorithm 1: p-Shrinkage Algorithm

Input : An $m \times N$ matrix A, initial guess $x^0 \in \mathbb{R}^N$, regularization parameter $\tau < \|A^T b\|_\infty^{2-p}$, tolerance β, maximum number of iterations M, and shrinkage parameter $p \in \mathbb{R}$.

Output: A sparse vector $\bar{x}$ with $\|A\bar{x} - b\|_2$ small.

```
for n = 0, 1, ..., M do
    x^{n+1} = R_{τ,p}(x^n + A^T b − A^T A x^n);
    if ||x^n − x^{n+1}||_2 ≤ β then
        break
    end
end
x̄ = x^{n+1};
```

4. NUMERICAL RESULTS

We have found that the p-thresholding scheme performs better than soft thresholding in many applications, with similar computation time. Below, we illustrate examples in compressive sensing (CS) and wavelet denoising. For CS, we have a (vectorized) sparse image x that we do not have access to. Instead we have access to measurements y of the image obtained with a random, normally distributed, underdetermined sensing matrix A, subject to some noise. That is, $y = Ax + \nu$ and we seek to recover x by minimizing $||Ax - y||_2^2 + 2\tau G_{p,\tau}(x)$. We find that when p-thresholding with $p < 1$ is used, fewer measurements are needed for sparse recovery than with the ISTA case of $p = 1$, or with the same number of measurements, better quality is obtained with $p < 1$.

In Fig. 3 we see results of examples using $p = 1/4$. In each case, our Gaussian A was scaled to have $\|A\|_2 = 10/11$ (as convergence requires this be less than 1). The noise ν has norm 10% of the norm of y. The first example shows reconstruction results for 2,744 measurements of an image with 5,025 pixels, of which 686 are nonzero. Using ISTA, the reconstruction SNR is 6.51 dB, while with our p-ISTA obtains 9.50 dB. The second example uses 1,145 measurements of a 4,662-pixel image with 229 nonzero pixels. The ISTA reconstruction has an SNR of 5.98 dB, while p-ISTA gives us 9.94 dB. Thus in each case we obtain a 3 dB improvement using our p-thresholding.

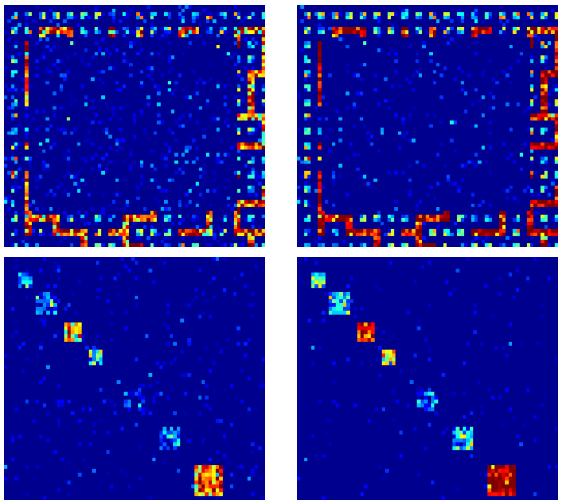

Fig. 3. Reconstructions with ISTA (left) and p-ISTA (right) with the same numbers of measurements, where $p = 1/4$.

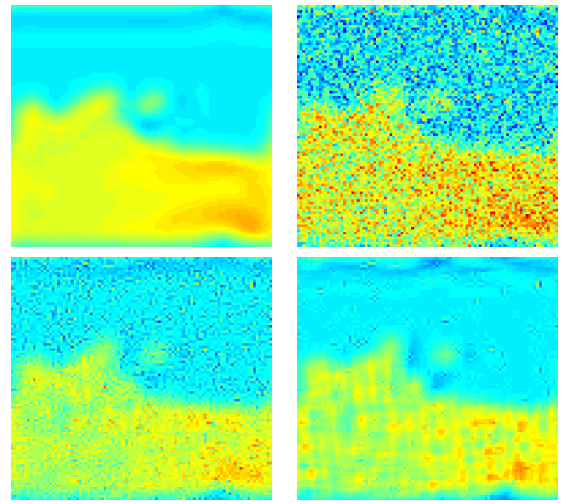

Fig. 4. Top: original and noisy images. Bottom: denoised images with ISTA (left) and p-ISTA (right).

Now we consider image denoising. In this example, we start with a 9,216-pixel satellite image, and then retain the largest 5% of the Daubechies-3 wavelet transform coefficients to construct our image x. We then obtain $y = x + \nu$ by adding Gaussian noise of norm 60% of the norm of x, giving an SNR of 1.38 dB. We minimize $||W^{-1}w - y||_2^2 + 2\tau G_{p,\tau}(w)$, where W is the db-3 wavelet transform. We find the minimizing $\bar{w}$ and construct $W^{-1}\bar{w}$ as the denoised image. Using ISTA, the reconstruction SNR is 6.27 dB, while using p-ISTA with $p = 3/4$ we obtain 8.60 dB, a significant improvement.

Another advantage of the new thresholding is that the p-ISTA algorithm can be run with a higher parameter τ than ISTA. This is advantageous since thresholding methods tend to be slow at small τ. Based on the first-order optimality conditions, it is straightforward to show the following:

Lemma 4.1. *For any $p \in \mathbb{R}$, $x = 0$ is a local minimizer of* (6) *iff $\tau \geq \|A^T b\|_\infty^{2-p}$.*

Thus, when $p < 1$, greater values of τ can be used to yield non-zero solutions than when $p = 1$.

Consider sparse vectors $x \in \mathbb{R}^{1500}$ with numbers of nonzeros in $\{75, 150, \ldots, 900\}$. In each case, we use the same 1000×1500 matrix A (constructed as before) to construct the right hand side $b = Ax + \nu$, with $\|\nu\|_2 = 0.1\|b\|_2$, and perform the inversion with ISTA and p-ISTA ($p = 0.65$) to find the x_τ such that $\|Ax_\tau - b\|_2 \approx \|\nu\|_2$. We loop over different values of τ, computing at each step the solution and the residual value. Once we have found the τ which leads to this noise matching residual level for each scheme, we run the algorithms until convergence: $\|x^n - x^{n-1}\|_2 < 10^{-5}$ and plot the number of iterations in Fig. 5. (The computation time per iteration was nearly the same for both methods.) We see that in some cases, p-ISTA takes somewhat longer to converge. We also tried incorporating the acceleration scheme of [12], which when applied to ISTA is known as FISTA. In this version, the thresholding is performed with x^n replaced by a linear combination y_n of two previous solutions, $y_n = x^n + ((t_n - 1)/t_{n+1})(x^n - x^{n-1})$, where $t_{n+1} = (1 + \sqrt{1 + 4t_n^2})/2$. This acceleration scheme is designed specifically for convex minimization, but as a heuristic it gives much faster convergence using p-thresholding as well. We will undertake to derive an acceleration approach designed for our p-thresholding in future work.

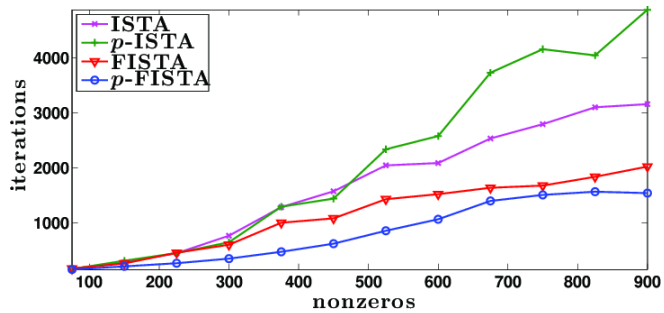

Fig. 5. Number of iterations until convergence criterion for increasing numbers of nonzeros in x.

5. RELATION TO PRIOR WORK

We presented a new generalized thresholding algorithm which uses a p-thresholding function, which reduces to soft thresholding when $p = 1$. For $p < 1$, the new thresholding function penalizes small coefficients more than soft thresholding, while applying less bias to the large coefficients. Our iterative p-thresholding algorithm has, in many cases, improved reconstruction performance compared with ISTA [2].

6. REFERENCES

[1] Rick Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *IEEE International Symposium on Biomedical Imaging*, 2009.

[2] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[3] David L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[4] Emmanuel J. Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, 2006.

[5] Rick Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[6] Rick Chartrand and Valentina Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[7] Sergey Voronin and Hugo J. Woerdeman, "A new iterative firm-thresholding algorithm for inverse problems with sparsity constraints," *Applied and Computational Harmonic Analysis*, 2012.

[8] Rick Chartrand, "Nonconvex regularization for shape preservation," in *IEEE International Conference on Image Processing (ICIP)*, 2007.

[9] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[10] David R. Hunter, Kenneth Lange, Departments Of Biomathematics, and Human Genetics, "A tutorial on mm algorithms," *Amer. Statist*, pp. 30–37, 2004.

[11] R. R. Meyer, "Sufficient conditions for the convergence of monotonic mathematical programming algorithms," *J. Comput. System Sci.*, pp. 108–121, 1976.

[12] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.