A SUBGRADIENT PROJECTION ALGORITHM FOR NON-DIFFERENTIABLE SIGNAL RECOVERY

Jian Luo and Patrick L. Combettes

Department of Electrical Engineering, City College and Graduate School,
City University of New York, New York, NY 10031, USA.

ABSTRACT

Many problems in signal recovery can be formulated as constrained convex minimization problems. In this paper, an adaptive level set subgradient projection algorithm is proposed to solve such problems with non-differentiable costs. Applications to minimum total variation signal restoration and denoising are demonstrated.

1. INTRODUCTION

Under consideration are signal recovery problems of the form

$$\text{Find } x^* \in Q \text{ such that } J(x^*) = \inf_{x \in Q} J(x) \triangleq \alpha^*, \quad (1)$$

where $J : \mathbb{R}^N \to \mathbb{R}$ is a convex function representing an optimality criterion and $Q \subset \mathbb{R}^N$ a compact convex set representing the constraints on the original signal. Practically speaking, the objective J allows the user to selectively pick a signal in the feasibility set Q. The reader is referred to [3, 13] for general background on signal recovery.

In recent years, interest has grown towards the use of non-differentiable costs in signal recovery, e.g., ℓ^1 cost in [7] and total variation cost in [15]. Other costs of interest include minimax costs, i.e., $J = \max_{i \in I} J_i$, where $(J_i)_{i \in I}$ is a family of convex functions. Solving (1) with such objectives poses specific problems which cannot be overcome by smooth optimization methods. Among the problems which may arise in using smooth optimization methods in nonsmooth problems, let us mention in particular the failure of convergence to a minimum and the lack of implementable stopping rules [8]. Nevertheless, the current practice in signal recovery is often to employ smooth algorithms and either ignore potential problems or simply "assume" that they will iterate in smooth regions.

Let us denote by $\mathrm{lev}_{\leq \alpha} J = \{x \in \mathbb{R}^N \mid J(x) \leq \alpha\}$ the lower level set of J at height $\alpha \in \mathbb{R}$. Then (1) can be written as

$$\text{Find } \; x^* \in S \triangleq Q \cap \mathrm{lev}_{\leq \alpha^*} J. \quad (2)$$

It should be noted that by virtue of the above assumptions, the solution set S is guaranteed to be nonempty, closed, and convex. If α^* is known, then (2) can be solved by Polyak's subgradient projection method [10, 12]: starting with $x_0 \in Q$, one passes from x_n to x_{n+1} by the rule

$$x_{n+1} = P_Q\left(x_n - \frac{(J(x_n) - \alpha^*)^+}{\|t_n\|^2} t_n\right), \quad (3)$$

where P_Q is the projector onto Q and t_n is a subgradient of J at x_n. It will be assumed throughout that $\inf_{x \in \mathbb{R}^N} J(x) < \alpha^*$ so that the subgradients at any $x \in Q$ are nonzero. In such instances, one essentially solves a convex feasibility problem with two sets. Furthermore, (3) appears as a special case of the algorithm proposed in [4] and every sequence it produces converges to a solution. Unfortunately, α^* is usually unknown and (3) must be modified into a level set method in which the true height α^* is adaptively estimated over the course of iterations [2, 5, 6].

The goal of this paper is to propose a general adaptive level set subgradient algorithm based on that of [6] for solving the signal recovery problems (1) with non-differentiable costs. The algorithm is presented in section 2 and analyzed in sections 3 and 4. Simulation results are presented in section 5 and section 6 concludes the paper with some remarks.

2. ALGORITHM

The principle of the algorithm is as follows. At iteration n, one picks a subgradient t_n of J at $x_n \in Q$ and forms the affine minorant

$$J_n : x \mapsto J(x_n) + \langle t_n \mid x - x_n \rangle \quad (4)$$

This work was supported by the National Science Foundation under grant MIP-9705504.

of J at x_n. Accordingly, the half-space $H_n = \text{lev}_{\leq \alpha_n} J_n$ stands for an outer approximation to the lower level set $\text{lev}_{\leq \alpha_n} J$. Now denote by P_{H_n} the projector onto H_n. Then most level set subgradient methods can be described by the recursion

$$\begin{aligned} x_{n+1} &= P_Q(P_{H_n}(x_n)) \\ &= P_Q\left(x_n - \frac{(J(x_n)-\alpha_n)^+}{\|t_n\|^2} t_n\right), \end{aligned} \quad (5)$$

where $(\alpha_n)_{n\geq 0}$ is constructed so as to approach α^* and $\tau^+ = \max\{0, \tau\}$. The construction of the sequence $(\alpha_n)_{n\geq 0}$ is detailed below.

Algorithm 1
Step 0. Fix $x_0 \in Q$, $\epsilon > 0$, and set $\alpha_0^- < \alpha^*$, $\alpha_{-1}^+ = +\infty$, and $n = 0$.
Step 1. Set $\alpha_n^+ = \min\{J(x_n), \alpha_{n-1}^+\}$.
Step 2. Set $\alpha_n = (\alpha_n^- + \alpha_n^+)/2$, $\Delta_n = \alpha_n^+ - \alpha_n^-$.
Step 3. If $\Delta_n \leq \epsilon$, terminate.
Step 4. If $Q \cap \text{lev}_{\leq \alpha_n} J = \emptyset$ is detected, go to step 6.
Step 5. Set $z_n = P_{H_n}(x_n)$, $x_{n+1} = P_Q(z_n)$, $\alpha_{n+1}^- = \alpha_n^-$, $n = n+1$, and go to step 1.
Step 6. Set $\alpha_{n+1}^- = \alpha_n$ and $\Delta_n = \alpha_n^+ - \alpha_{n+1}^-$.
Step 7. If $\Delta_n \leq \epsilon$, terminate. Otherwise, set $x_{n+1} = x_n$, $\alpha_{n+1}^+ = \alpha_n^+$, $n = n+1$, and go to step 2.

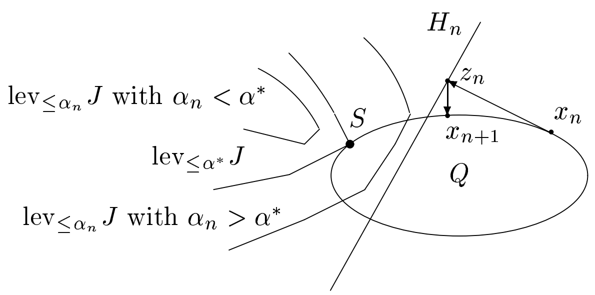

Figure 1: Alternating projection scheme (5).

As illustrated in Figure 1, step 5 consists a projection onto the approximating half-space H_n followed by a projection onto the constraint set Q. It is noted that the iterates $(x_n)_{n\geq 0}$ lie in Q.

3. ASYMPTOTIC PROPERTIES

In this section, ϵ is the tolerance fixed at step 0 of Algorithm 1. By construction, we have

$$(\forall n) \;\; \alpha_{n-1}^+ \geq \alpha_n^+ = \min_{0\leq i\leq n} J(x_i) \geq \alpha^*. \quad (6)$$

On the other hand, for every iteration n, $\alpha_n^- \in [\alpha_0^-, \alpha^*[$ and it is updated by the rule

$$\alpha_{n+1}^- = \begin{cases} \alpha_n^- & \text{if } Q \cap \text{lev}_{\leq \alpha_n} J \neq \emptyset \\ \alpha_n^- & \text{if } Q \cap \text{lev}_{\leq \alpha_n} J = \emptyset \text{ is not detected} \\ \alpha_n & \text{if } Q \cap \text{lev}_{\leq \alpha_n} J = \emptyset \text{ is detected.} \end{cases} \quad (7)$$

Note that the update $\alpha_{n+1}^- = \alpha_n^- + (\alpha_n^+ - \alpha_n^-)/2$ takes place only if infeasibility is detected at step 4. Now define

$$(\forall n) \;\; N_n = \{k \in \mathbb{N} \mid \alpha_k^- = \alpha_n^-\}. \quad (8)$$

Then, in view of step 2, the (possibly finite) sequence $(\alpha_k)_{k\in N_n}$ is nonincreasing.

Proposition 1 [9] $(\forall n)$ $(\exists\, k \in N_n)$ $\alpha_k < \alpha^*$. Now suppose that

$$(\forall n)(\exists\, k \in N_n)\; Q \cap \text{lev}_{\leq \alpha_k} J = \emptyset \text{ is detected.} \quad (9)$$

Then

$$(\exists\, n \in \mathbb{N}) \;\; \begin{cases} \alpha^* - \alpha_n^- \leq \epsilon \\ (\exists\, k \in N_n)\; \alpha_k^+ \leq \alpha^* + \epsilon. \end{cases} \quad (10)$$

From (6) and Proposition 1, we deduce the following convergence result.

Theorem 1 [9] Suppose that infeasibility at step 4 can be detected. Then Algorithm 1 generates a point x_n such that $|J(x_n) - \alpha^*| \leq \epsilon$.

As will be seen shortly, infeasibility can indeed be detected. Therefore, the above theorem states that Algorithm 1 produces a signal that achieves any preset tolerance value on the optimal value of the objective.

4. IMPLEMENTATION

The implementation of Algorithm 1 is straightforward except for infeasibility detection at step 4. In this section, we provide explicit schemes for its implementation.

At every iteration n, the sequence $(\alpha_k)_{k\in N_n}$ is nonincreasing and, by Proposition 1,

$$(\exists k \in N_n)\; Q \cap \text{lev}_{\leq \alpha_k} J = \emptyset. \quad (11)$$

Unfortunately, such a k cannot be identified directly unless the sets Q and $\text{lev}_{\leq \alpha_n} J$ are simple. However, if one knows *a priori* the diameter κ of Q or an upper bound γ on $d(x_0, S)$, the following proposition provides sufficient conditions for detecting infeasibility in (11).

Proposition 2 [9] At iteration n, define for every $k \in N_n$ $\rho_k^H = \|x_k - z_k\|^2$, $\rho_k^Q = \|z_k - x_{k+1}\|^2$, and $\rho_k = \rho_k^H + \rho_k^Q$. Then, for any $(n_1, n_2) \in N_n^2$ such that $n_1 \leq n_2$, $Q \cap \mathrm{lev}_{\leq \alpha_{n_2}} J = \emptyset$ if any of the following holds:

(i) $\sum_{k=n_1}^{n_2} \rho_k > \kappa^2$;

(ii) $\sum_{k=n_1}^{n_2} \rho_k > \gamma^2$;

(iii) $\sum_{k=n_1}^{n_2} \rho_k > 2\gamma\|x_{n_1} - x_{n_2+1}\| - \|x_{n_1} - x_{n_2+1}\|^2$.

Although the conditions stated in Proposition 2 are not necessary, they can be used to detect (11) by virtue of the following proposition.

Proposition 3 [9] If N_n is infinite then $\sum_{k \in N_n} \rho_k^H = +\infty$ and $\sum_{k \in N_n} \rho_k = +\infty$.

Note that if (11) occurs at iteration k and is not detected at that iteration, it will occur again at iteration $k+1$. Now, if infeasibility is not detected at any iteration beyond k, N_n will be infinite as a consequence of (7). However, by Proposition 3, as N_n grows the sums in Proposition 2(i)-(iii) will become arbitrarily large and therefore infeasibility will be detected. Since N_n can easily be constructed, the conditions given in Proposition 2 for detecting infeasibility are implementable.

When using (i) or (ii) in Proposition 2 to detect infeasibility at step 4, Algorithm 1 turns out to be very similar to that proposed in [5]. The difference is that we use $\sum_{k \in N_n} \rho_k$, as opposed to $\sum_{k \in N_n} \rho_k^H$ in [5], which allows us to detect infeasibility sooner.

5. APPLICATION TO MINIMUM TOTAL VARIATION SIGNAL RECOVERY

5.1. Generalities

The total variation of a 2-dimensional analog signal x defined on an open set $\Omega \subset \mathbb{R}^2$ is

$$J_{TV}(\mathsf{x}) = \int_\Omega |\nabla \mathsf{x}(\omega)| d\omega, \tag{12}$$

where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^2$ [1]. The discrete counterpart of (12) can be defined as

$$J_{TV}(x) = \sum_{i,j} \sqrt{(x^{i,j} - x^{i,j+1})^2 + (x^{i,j} - x^{i+1,j})^2}, \tag{13}$$

where $x^{i,j}$ denotes the (i,j) component of $x \in \mathbb{R}^{N \times N}$. It is easy to show that J_{TV} as defined by (13) is a non-differentiable convex function on $\mathbb{R}^{N \times N}$. Unlike energy or Laplacian costs, such costs do not impose a strong smoothness condition and they are therefore of interest in the recovery of sharp signals. They have in particular been applied with some success to the recovery of blocky signals (see [15] and the references therein).

Consider the standard linear degradation model

$$y = Tx + u, \tag{14}$$

where y and u are respectively the observed signal and the additive noise, and T an $N \times N$ matrix. Under certain statistical hypotheses on the noise vector u, we can define the constraint set [3, 14]

$$Q = \{x \in \mathbb{R}^N \mid \|Tx - y\|^2 \leq \delta\}. \tag{15}$$

One is then led to the problem formulation (1), consisting of finding a signal x^* of minimum total variation and consistent with the observed data and the noise information, i.e,

$$\text{Find } x^* \in Q \text{ such that } J_{TV}(x^*) = \inf_{x \in Q} J_{TV}(x). \tag{16}$$

This problem can be solved in the sense of Theorem 1 by Algorithm 1.

In [11], the method used to solve (16) is akin to a gradient projection method in which iterates are perturbed to avoid points of non-differentiability. This approach leads to a straightforward numerical implementation but it lacks a sound mathematical basis, and exact convergence properties such as those stated in Theorem 1 do not seem achievable.

In [15], (16) is reformulated as a penalized problem of the form (actually, J_{TV} was slightly modified to avoid non-differentiability issues)

$$\text{Find } \arg \min_{x \in \mathbb{R}^{N \times N}} \{J_{TV}(x) + \lambda(\|Tx - y\|^2 - \delta)\}. \tag{17}$$

Of course, problems (16) and (17) are not equivalent, unless λ is a Kuhn-Tucker coefficient for the former. In [15], the choice of λ is guided by heuristic rules and therefore the properties of the solutions to (17) are not clear.

5.2. Example of Signal Restoration

As illustrated in the following example, total variation works quite well as a minimizing criterion for restoring "blocky" signals even when the level of the noise is high. The degraded signal y shown in Figure 4 is obtained by convolving the point spread function shown in Figure 2 with the original signal x shown in Figure 3 and addition of zero mean white Gaussian. The signal-to-noise ratio is 14.94dB.

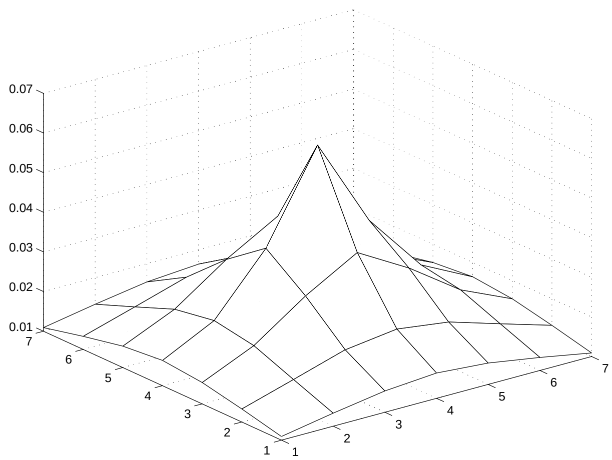

Figure 2: Point spread function.

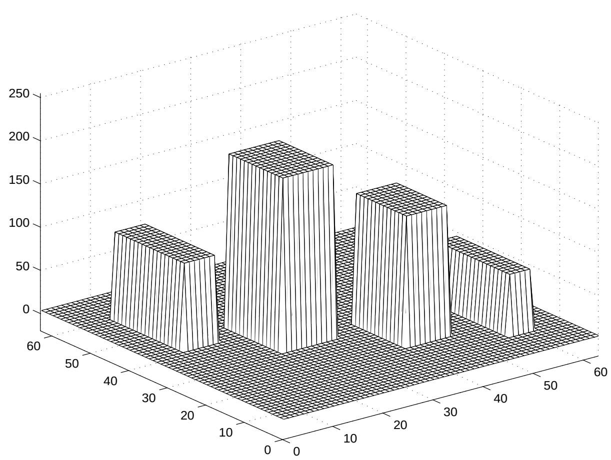

Figure 3: Original signal.

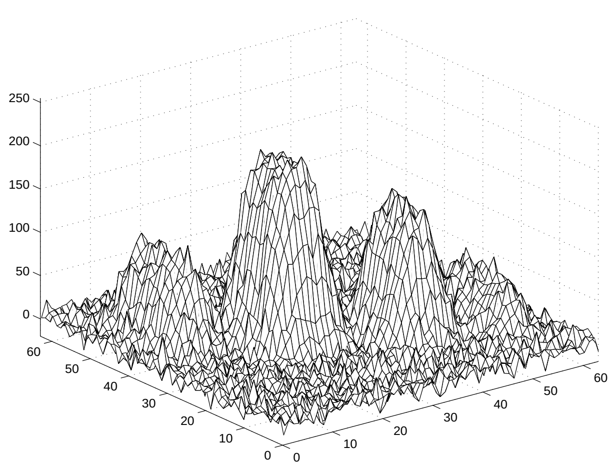

Figure 4: Degraded signal.

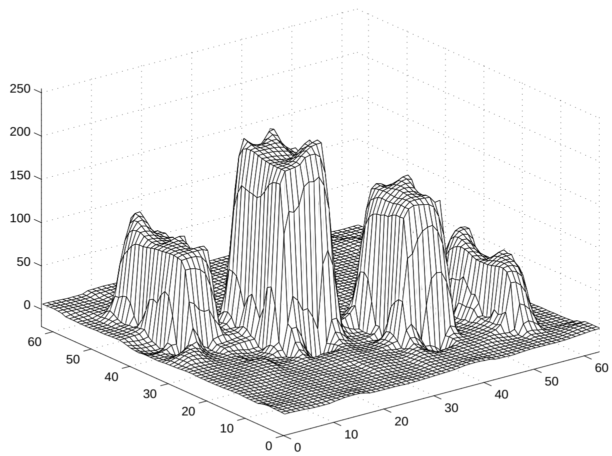

Figure 5: Restored signal.

Algorithm 1 generated the signal shown in Figure 5 and it can be seen that the "block" features of the original signal are fairly well restored.

5.3. Example of Signal Denoising

The noisy signal shown in Figure 6 is obtained by adding zero mean white Gaussian noise to the original signal x shown in Figure 3. The signal-to-noise ratio is 9.57dB. Here T is the identity matrix and the constraint set Q in (15) becomes a ball. The denoised signal produced by Algorithm 1 is shown in Figure 7.

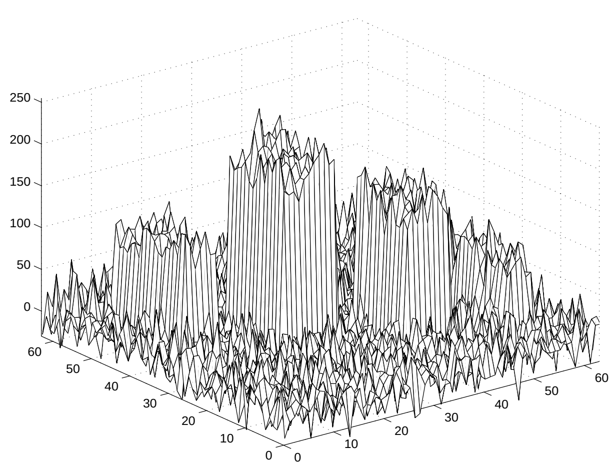

Figure 6: Noisy signal.

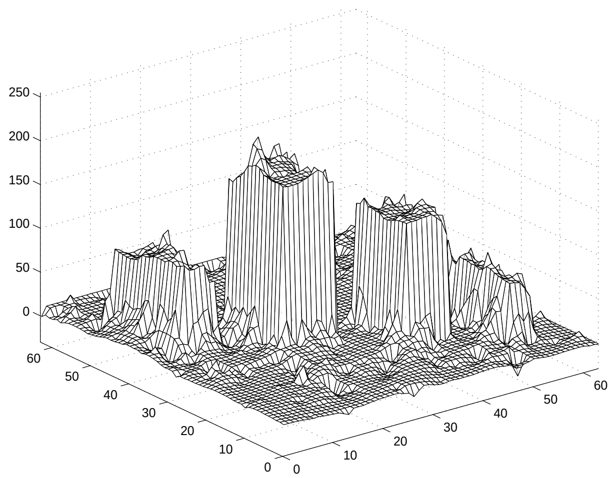

Figure 7: Denoised signal.

6. CONCLUSION

A general algorithm has been proposed for solving constrained convex signal recovery problems with non-differentiable costs. The algorithm has been shown to produce a signal that achieves any preset tolerance level on the optimal value of the objective. Unlike many methods currently in use in nonsmooth signal recovery, it produces a solution with known properties. As regards to its asymptotic behavior, let us note that as α_n^- approaches the optimal value α^* the number of iterations required to detect infeasibility increases. Therefore it is important to assign to the tolerance parameter ϵ a realistic value; too small a value will increase the number of iterations with no practical improvement on the solution.

As with any projected (sub)gradient method, the constraint set should be simple enough so that the projection onto it can be implemented at reasonably low cost. Extending Theorem 1 to cases in which $P_Q(z_n)$ is approximately computed at step 5 would make the incorporation of a wide range of constraints numerically viable. Another direction of research is to explore the possibility of establishing tighter bounds than those proposed in Proposition 2 in specific problems. This would make the algorithm faster by allowing for early infeasibility detection at step 4.

7. REFERENCES

[1] R. Acar and C. R. Vogel, "Analysis of bounded variation penalty methods for ill-posed problems," *Inverse Problems,* vol. 10, pp. 1217-1229, 1994.

[2] E. Allen, R. Helgason, J. Kennington, and B. Shetty, "A generalization of Polyak's convergence result for subgradient optimization," *Mathematical Programming,* vol. 37, pp. 309-317, 1987.

[3] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics,* vol. 95, pp. 155-270. New York: Academic, 1996.

[4] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Transactions on Image Processing,* vol. 6, pp. 493-506, 1997.

[5] S. Kim, H. Ahn, and S. Cho, "Variable target value subgradient method," *Mathematical Programming,* vol. 49, pp. 359-369, 1991.

[6] K. C. Kiwiel, "The efficiency of subgradient projection methods for convex optimization, Parts I and II", *SIAM Journal on Control and Optimization,* vol. 34, pp. 660-676 and pp. 677-697, 1996.

[7] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image reconstruction," *IEEE Transactions on Image Processing,* vol. 5, pp. 987-995, 1996.

[8] C. Lemaréchal, "Nondifferentiable optimization," in *Handbooks in Operations Research and Management Science,* vol. 1, pp. 529-572. New York: North-Holland, 1989.

[9] J. Luo, *Non-Differentiable Signal Recovery,* Ph.D. dissertation, City University of New York, in progress.

[10] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics,* vol. 9, pp. 14-29, 1969.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259-268, 1992.

[12] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions.* New York: Springer-Verlag, 1985.

[13] H. Stark (Editor), *Image Recovery: Theory and Application.* San Diego, CA: Academic Press, 1987.

[14] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 32, pp. 201-212, 1984.

[15] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing,* vol. 7, pp. 813-824, 1998.