

SPLIT CONVEX MINIMIZATION ALGORITHM FOR SIGNAL RECOVERY

Patrick L. Combettes

Jean-Christophe Pesquet

UPMC Université Paris 06
Laboratoire Jacques-Louis Lions – UMR 7598
75005 Paris, France

Université Paris-Est
Institut Gaspard Monge – UMR 8049
77454 Marne la Vallée Cedex 2, France

ABSTRACT

A broad range of signal recovery problems can be abstracted into the problem of minimizing the sum of several convex functions in a Hilbert space. We propose a proximal decomposition algorithm which, under mild conditions, provides a solution to such a problem. A significant improvement over the methods currently in use in the area of signal recovery is that it is not limited to two nondifferentiable functions. An application to image restoration is demonstrated.

Index Terms— convex optimization methods, inverse problems, parallel algorithm, signal restoration, variational methods, wavelet transforms.

1. INTRODUCTION

A broad range of signal recovery problems can be formulated as decomposed optimization problems of the form

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^m f_i(x), \quad (1.1)$$

where $(\mathcal{H}, \|\cdot\|)$ is a real Hilbert space and $(f_i)_{1 \leq i \leq m}$ are proper lower semicontinuous convex functions from \mathcal{H} to $]-\infty, +\infty]$. In this variational formulation, each potential function f_i may represent a prior constraint on the ideal signal \bar{x} or on the data acquisition model. The purpose of this paper is to propose a decomposition method that, under rather general conditions, will provide solutions to (1.1).

To place our investigation in perspective, let us review some important special cases of (1.1) for which globally convergent numerical methods are available. If the functions $(f_i)_{1 \leq i \leq m}$ are the indicator functions (see (2.1)) of convex constraint sets $(C_i)_{1 \leq i \leq m}$, (1.1) reduces to the convex feasibility problem [16], which can be solved by projection techniques [6]. When the sets $(C_i)_{1 \leq i \leq m}$ are based on inaccurate information, the feasibility set $\bigcap_{i=1}^m C_i$ may be empty. An approximate solution can be obtained by setting, for every $i \in \{1, \dots, m\}$, $f_i = \omega_i d_{C_i}^2$, where d_{C_i} is the distance function to C_i (see Section 2) and where $\omega_i \in]0, 1]$ [5];

This work was supported by the Agence Nationale de la Recherche under grant ANR-05-MMSA-0014-01.

for finite-dimensional variants based on Bregman distances, see [2]. If $(f_i)_{1 \leq i \leq m-1}$ are the indicator functions of convex sets $(C_i)_{1 \leq i \leq m-1}$ and $f_m: x \mapsto \|x - r\|^2$ for some $r \in \mathcal{H}$, then (1.1) reduces to a best approximation problem [7]. In [11], the special instance of (1.1) in which $m = 2$ and f_2 is Lipschitz-differentiable on \mathcal{H} is shown to cover a variety of seemingly unrelated signal recovery formulations, and a forward-backward splitting algorithm is proposed to solve this problem. This 2-function framework can be extended to (1.1) under the restriction that $(f_i)_{2 \leq i \leq m}$ be Lipschitz-differentiable. Finally, the setting of [9] corresponds to $m = 2$ in (1.1), without any differentiability assumption. The algorithm adopted in [9] is based on the Douglas-Rachford splitting method, namely

$$\begin{cases} y_{n+\frac{1}{2}} = \text{prox}_{\gamma f_2} y_n + a_n \\ y_{n+1} = y_n + \lambda_n \left(\text{prox}_{\gamma f_1} (2y_{n+\frac{1}{2}} - y_n) + b_n - y_{n+\frac{1}{2}} \right), \end{cases} \quad (1.2)$$

where $\lambda_n \in]0, 2[$, $\gamma > 0$, and a_n and b_n model tolerances in the implementation of the proximity operators (see (2.2)). Under suitable assumptions, $(y_n)_{n \in \mathbb{N}}$ converges weakly to a point $y \in \mathcal{H}$ and $\text{prox}_{\gamma f_2} y$ minimizes $f_1 + f_2$ [9].

Some important scenarios are not covered by the above settings, namely the formulations of type (1.1) with three or more potentials, at least two of which are non differentiable, e.g., ℓ^1 norm, total variation, distance functions, max functions, support functions, etc. Our main objective is to propose an algorithm to overcome this limitation. In Section 2, we provide some background on convex analysis and proximity operators. In Section 3, we introduce our algorithm and provide conditions for its convergence. An application to image restoration is detailed in Section 4.

2. NOTATION AND BACKGROUND

Let C be a nonempty convex subset of \mathcal{H} . The indicator function of C is

$$\iota_C: x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C, \end{cases} \quad (2.1)$$

and its distance function is $d_C: \mathcal{H} \rightarrow [0, +\infty[: x \mapsto \inf_{y \in C} \|x - y\|$. In addition, if C is closed, the projection of a point x in \mathcal{H} onto C is the unique point $P_C x$ in C such that $\|x - P_C x\| = d_C(x)$.

The domain of a function $f: \mathcal{H} \rightarrow]-\infty, +\infty]$ is $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$. The class of lower semicontinuous convex functions from \mathcal{H} to $]-\infty, +\infty]$ which are proper (i.e., with nonempty domain) is denoted by $\Gamma_0(\mathcal{H})$. The proximity operator of a function $f \in \Gamma_0(\mathcal{H})$ is the operator $\text{prox}_f: \mathcal{H} \rightarrow \mathcal{H}$ which uniquely maps every $x \in \mathcal{H}$ to

$$\text{prox}_f x = \underset{y \in \mathcal{H}}{\text{argmin}} \quad f(y) + \frac{1}{2} \|x - y\|^2. \quad (2.2)$$

Closed-form formulas for various proximity operators are provided in [4, 9, 11]. The following two properties will be required.

Proposition 2.1 [9, Proposition 11] *Let \mathcal{G} be a real Hilbert space, let $f \in \Gamma_0(\mathcal{G})$, and let $L: \mathcal{H} \rightarrow \mathcal{G}$ be a bounded linear operator with closed range $\text{ran } L$ and such that $L \circ L^* = \kappa \text{Id}$, for some $\kappa \in]0, +\infty[$. Suppose that $0 \in \text{int}(\text{dom } f - \text{ran } L)$. Then $f \circ L \in \Gamma_0(\mathcal{H})$ and $\text{prox}_{f \circ L} = \text{Id} + \kappa^{-1} L^* \circ (\text{prox}_{\kappa f} - \text{Id}) \circ L$.*

Proposition 2.2 [10] *Let $(\mathcal{G}, \|\cdot\|)$ be a real Hilbert space, let $L: \mathcal{H} \rightarrow \mathcal{G}$ be linear and bounded, let $z \in \mathcal{G}$, let $\gamma \in]0, +\infty[$, and set $f = \gamma \|L \cdot - z\|^2 / 2$. Then $f \in \Gamma_0(\mathcal{H})$ and*

$$(\forall x \in \mathcal{H}) \quad \text{prox}_f x = (\text{Id} + \gamma L^* L)^{-1} (x + \gamma L^* z). \quad (2.3)$$

3. ALGORITHM AND CONVERGENCE

We propose the following proximal method to solve (1.1).

Algorithm 3.1 For every $i \in \{1, \dots, m\}$, let $(a_{i,n})_{n \in \mathbb{N}}$ be a sequence in \mathcal{H} . A sequence $(x_n)_{n \in \mathbb{N}}$ is generated by the following routine.

Initialization

$$\left\{ \begin{array}{l} \gamma \in]0, +\infty[\\ (\omega_i)_{1 \leq i \leq m} \in]0, 1]^m \text{ satisfy } \sum_{i=1}^m \omega_i = 1 \\ (y_{i,0})_{1 \leq i \leq m} \in \mathcal{H}^m \\ x_0 = \sum_{i=1}^m \omega_i y_{i,0} \end{array} \right.$$

For $n = 0, 1, \dots$

$$\left\{ \begin{array}{l} \text{For } i = 1, \dots, m \\ \quad \left[\begin{array}{l} p_{i,n} = \text{prox}_{\gamma f_i / \omega_i} y_{i,n} + a_{i,n} \\ p_n = \sum_{i=1}^m \omega_i p_{i,n} \\ \lambda_n \in]0, 2[\end{array} \right. \\ \text{For } i = 1, \dots, m \\ \quad \left[\begin{array}{l} y_{i,n+1} = y_{i,n} + \lambda_n (2p_n - x_n - p_{i,n}) \\ x_{n+1} = x_n + \lambda_n (p_n - x_n) \end{array} \right. \end{array} \right. \quad (3.1)$$

In this splitting algorithm, each function f_i is used separately by means of its own proximity operator. Thus, Algorithm 3.1 has a parallel structure in that, at iteration n , the proximal vectors $(p_{i,n})_{1 \leq i \leq m}$, as well as the auxiliary vectors $(y_{i,n})_{1 \leq i \leq m}$, can be computed simultaneously. Another feature of the algorithm is that some error $a_{i,n}$ is tolerated in the computation of the i th proximity operator. Its convergence is secured by the following result.

Theorem 3.2 [10] *Let G be the set of solutions to (1.1) and let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 3.1 under the following assumptions.*

- (i) $\lim_{\|x\| \rightarrow +\infty} f_1(x) + \dots + f_m(x) = +\infty$.
- (ii) $\text{dom } f_1 \cap \bigcap_{i=2}^m \text{int } \text{dom } f_i \neq \emptyset$.
- (iii) $\sum_{n \in \mathbb{N}} \lambda_n (2 - \lambda_n) = +\infty$.
- (iv) $(\forall i \in \{1, \dots, m\}) \sum_{n \in \mathbb{N}} \lambda_n \|a_{i,n}\| < +\infty$.

Then $G \neq \emptyset$ and $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in G .

Remark 3.3

- If \mathcal{H} is finite-dimensional, condition (ii) can be replaced by $\bigcap_{i=1}^m \text{ri } \text{dom } f_i \neq \emptyset$, where ri denotes the relative interior [10].
- When $m = 2$, Algorithm 3.1 does not revert to the standard Douglas-Rachford iteration (1.2). Actually, even in this case, it seems better to use the former to the extent that, as seen in Theorem 3.2, it produces directly a sequence that converges weakly to a minimizer of $f_1 + f_2$, whereas the latter does not.
- A practical limitation of the algorithm is that one should be able to implement, to within some controllable error, the proximity operators of each function.

4. APPLICATION TO IMAGE RESTORATION

In image recovery, variational formulations involving total variation [3, 14] or sparsity promoting potentials [1, 8, 12, 13] are popular. The objective of the present experiment is to show that it is possible to employ more sophisticated, hybrid potentials.

In order to simplify our presentation, we place ourselves in the Hilbert space \mathcal{G} of periodic discrete images $y = (\eta_{k,l})_{(k,l) \in \mathbb{Z}^2}$ with horizontal and vertical periods equal to N ($N = 512$), endowed with the standard Euclidean norm. As usual, images of size $N \times N$ are viewed as elements of this space through periodization. The original 8-bit satellite image $\bar{y} \in \mathcal{G}$ displayed in Figure 1 is degraded through the linear model $z = L\bar{y} + w$, where L is the two-dimensional periodic convolution operator with a

7×7 uniform kernel, and w is a realization of a periodic zero-mean white Gaussian noise. The resulting degraded image $z \in \mathcal{G}$ is shown in Figure 2. The blurred image-to-noise ratio is $20 \log_{10}(\|L\bar{y}\|/\|w\|) = 20.71$ dB and the relative quadratic error with respect to the original image is $20 \log_{10}(\|z - \bar{y}\|/\|\bar{y}\|) = -12.02$ dB.

In the spirit of a number of recent investigations (see [4] and the references therein), we use a tight frame representation of the images under consideration. This representation is defined through a synthesis operator F^* , which is a linear operator from $\mathcal{H} = \mathbb{R}^K$ to \mathcal{G} (with $K \geq N^2$) such that $F^* \circ F = \kappa \text{Id}$, for some $\kappa \in]0, +\infty[$. Thus, the original image can be written as $\bar{y} = F^* \bar{x}$, where $\bar{x} \in \mathcal{H}$ is a vector of frame coefficients to be estimated. The rationale behind this approach is that, by appropriately choosing the frame, a sparse representation \bar{x} of \bar{y} can be achieved.

The restoration problem is posed in the frame coefficient space \mathcal{H} . We use the constraint set imposing the range of the pixel values of the original image \bar{y} , namely

$$C = \{x \in \mathcal{H} \mid F^* x \in D\}, \quad (4.1)$$

where

$$D = \{y \in \mathcal{G} \mid (\forall (k, l) \in \{0, \dots, N-1\}^2) \eta_{k,l} \in [0, 255]\}, \quad (4.2)$$

as well as three potentials. The first potential is the standard least-squares data fidelity term $x \mapsto \|LF^* x - z\|^2$. The second potential is the ℓ^1 norm, which aims at promoting a sparse frame representation [4, 12, 15]. Finally, the third potential is the discrete total variation tv , which aims at preserving piecewise smooth areas and sharp edges [3, 14]. Using the notation $(\eta_{k,l})_{(k,l) \in \mathbb{Z}^2}^\top = (\eta_{l,k})_{(k,l) \in \mathbb{Z}^2}$, the discrete total variation of $y \in \mathcal{G}$ is defined as

$$\text{tv}(y) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \varrho_{k,l} (\nabla_1 y, (\nabla_1 (y^\top))^\top), \quad (4.3)$$

where $\nabla_1 : \mathcal{G} \rightarrow \mathbb{R}^{N \times N}$ is a discrete vertical gradient operator and where, for every k and l in $\{0, \dots, N-1\}$,

$$\begin{aligned} \varrho_{k,l} : \left([\nu_{a,b}]_{0 \leq a, b \leq N-1}, [\tilde{\nu}_{a,b}]_{0 \leq a, b \leq N-1} \right) \\ \mapsto \sqrt{|\nu_{k,l}|^2 + |\tilde{\nu}_{k,l}|^2}. \end{aligned} \quad (4.4)$$

A common choice for the gradient operator is $\nabla_1 : y \mapsto [\eta_{k+1,l} - \eta_{k,l}]_{0 \leq k, l \leq N-1}$. As is customary in image processing, we adopt here a horizontally smoothed version of this operator, namely,

$$\nabla_1 : y \mapsto \frac{1}{2} [\eta_{k+1,l+1} - \eta_{k,l+1} + \eta_{k+1,l} - \eta_{k,l}]_{0 \leq k, l \leq N-1}. \quad (4.5)$$

We thus arrive at a variational formulation of the form (1.1), namely

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \iota_C(x) + \|LF^* x - z\|^2 + \alpha \|x\|_{\ell^1} + \beta \text{tv}(F^* x), \quad (4.6)$$

where α and β are in $]0, +\infty[$. We are not aware of an efficient splitting method that produces sequences with guaranteed convergence to a solution to this problem. However, we shall see that it can be solved by the proposed method. Since C is bounded, condition (i) in Theorem 3.2 is satisfied. In addition, it is clear that condition (ii) in Theorem 3.2 also holds. Indeed, all the potentials in (4.6) have full domain, except ι_C .

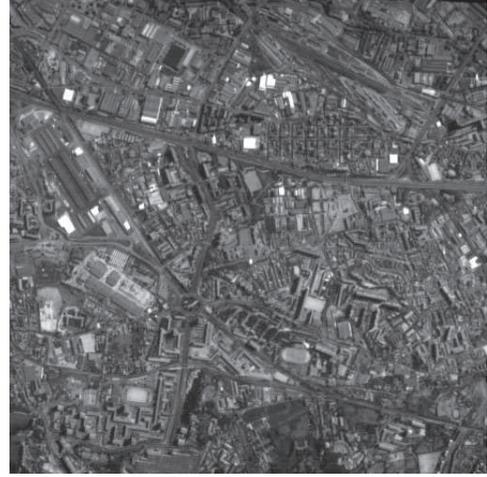


Fig. 1. Original image.

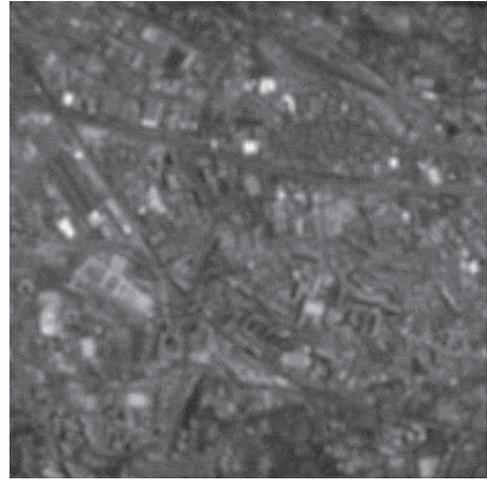


Fig. 2. Degraded image.

Although (4.6) assumes the form of (1.1), it is not directly exploitable by Algorithm 3.1 because the proximity operator of $\text{tv} \circ F^*$ cannot be computed explicitly. To circumvent this numerical hurdle, the total variation potential (4.3) is split in four terms and (4.6) is rewritten as

$$\underset{x \in C}{\text{minimize}} \quad \|LF^* x - z\|^2 + \alpha \|x\|_{\ell^1} + \beta \sum_{i=0}^3 \text{tv}_i(F^* x) \quad (4.7)$$

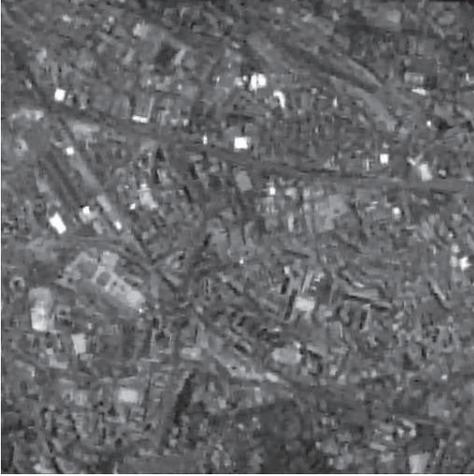


Fig. 3. Image restored by (4.7), using 350 iterations of Algorithm 3.1 with $\gamma = 150$.

where, for every q and r in $\{0, 1\}$ and every $y \in \mathcal{G}$,

$$\text{tv}_{q+2r}(y) = \sum_{k=0}^{N/2-1} \sum_{l=0}^{N/2-1} \varrho_{2k+q, 2l+r} (\nabla_1 y, (\nabla_1 (y^\top))^\top). \quad (4.8)$$

Problem (4.7) is a specialization of (1.1), in which $m = 7$, $f_1 = \iota_C$, $f_2 = \|LF^* \cdot -z\|^2$, $f_3 = \alpha \|\cdot\|_{\ell^1}$, and $f_{i+4} = \beta \text{tv}_i \circ F^*$ for $i \in \{0, 1, 2, 3\}$. Let us stress that only the function f_2 is differentiable. To implement Algorithm 3.1, we need the expressions of the proximity operators of $(f_i)_{1 \leq i \leq 7}$. The proximity operator of f_1 can be calculated by first observing that the projection onto the set D of (4.2) is explicit, and by then applying Proposition 2.1. On the other hand, the proximity operator of f_2 can be derived from Proposition 2.2 using a frequency domain implementation, and by again invoking Proposition 2.1. Next, the proximity operator of f_3 can be found in [11, Example 2.20]. Finally, the operators $(\text{prox}_{f_i})_{4 \leq i \leq 7}$ are provided by [10, Proposition 4.1].

In (4.7), we employ a tight frame ($\kappa = 4$) resulting from the concatenation of four shifted separable dyadic orthonormal wavelet decompositions carried out over 4 resolution levels. The shift parameters are $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$. In addition, symlet filters of length 8 are used. The parameters α and β have been adjusted so as to minimize the error with respect to the original image \bar{y} . The restored image we obtain is displayed in Figure 3. It achieves a relative mean-square error with respect to \bar{y} of -14.82 dB. For comparison, the result obtained without the total variation potential in (4.7) yields an error of -14.06 dB, and the result obtained without the ℓ^1 potential in (4.7) yields an error of -13.70 dB. This shows the advantage of combining an ℓ^1 potential and a total variation potential.

5. REFERENCES

- [1] J. M. Bioucas-Dias and M. A. Figueiredo, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration, *IEEE Trans. Image Process.*, vol. 16, pp. 2992–3004, 2007.
- [2] C. L. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Ann. Oper. Res.*, vol. 105, pp. 77–98, 2001.
- [3] A. Chambolle and P.-L. Lions, Image recovery via total variation minimization and related problems, *Numer. Math.*, vol. 76, pp. 167–188, 1997.
- [4] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, A variational formulation for frame-based inverse problems, *Inverse Problems*, vol. 23, pp. 1495–1518, 2007.
- [5] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Trans. Signal Process.*, vol. 42, pp. 2955–2966, 1994.
- [6] P. L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Trans. Image Process.*, vol. 6, pp. 493–506, 1997.
- [7] P. L. Combettes, A block-iterative surrogate constraint splitting method for quadratic signal recovery, *IEEE Trans. Signal Process.*, vol. 51, pp. 1771–1782, 2003.
- [8] P. L. Combettes and J.-C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, *SIAM J. Optim.*, vol. 18, pp. 1351–1376, 2007.
- [9] P. L. Combettes and J.-C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, *IEEE J. Selected Topics Signal Process.*, vol. 1, pp. 564–574, 2007.
- [10] P. L. Combettes and J.-C. Pesquet, A proximal decomposition method for solving convex variational inverse problems, *Inverse Problems*, vol. 24, no. 6, 2008.
- [11] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.
- [12] I. Daubechies, M. Defrise, and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.
- [13] I. Daubechies, G. Teschke, and L. Vese, Iteratively solving linear inverse problems under general convex constraints, *Inverse Problems and Imaging*, vol. 1, pp. 29–46, 2007.
- [14] L. I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, vol. 60, pp. 259–268, 1992.
- [15] J. A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inform. Theory*, vol. 52, pp. 1030–1051, 2006.
- [16] D. C. Youla and H. Webb, Image restoration by the method of convex projections: Part I – theory, *IEEE Trans. Medical Imaging*, vol. 1, pp. 81–94, 1982.