Alternating proximal algorithm for blind image recovery

Jérôme Bolte, Patrick Louis Combettes, Jean-Christophe Pesquet

ALTERNATING PROXIMAL ALGORITHM FOR BLIND IMAGE RECOVERY

J. Bolte,[1] *P. L. Combettes,*[2] *and J.-C. Pesquet,*[3]

[1]UPMC Université Paris 06, Équipe Combinatoire et Optimisation, UMR CNRS 7090, 75005 Paris, France
[2]UPMC Université Paris 06, Laboratoire Jacques-Louis Lions, UMR CNRS 7598, 75005 Paris, France
[3]Université Paris-Est, Lab. d'Informatique Gaspard Monge, UMR CNRS 8049, 77454 Marne la Vallée Cedex 2, France

ABSTRACT

We consider a variational formulation of blind image recovery problems. A novel iterative proximal algorithm is proposed to solve the associated nonconvex minimization problem. Under suitable assumptions, this algorithm is shown to have better convergence properties than standard alternating minimization techniques. The objective function includes a smooth convex data fidelity term and nonsmooth convex regularization terms modeling prior information on the data and on the unknown linear degradation operator. A novelty of our approach is to bring into play recent nonsmooth analysis results. The pertinence of the proposed method is illustrated in an image restoration example.

Index Terms— Blind restoration, blind reconstruction, proximal methods, nonlinear optimization, wavelets

1. INTRODUCTION

Blind restoration and reconstruction are challenging problems in image processing [3, 4, 9]. Variational approaches to these problems are often based on alternating minimization strategies which, in spite of their practical usefulness, offer in general few theoretical guarantees of convergence. In this paper, we propose a novel proximal alternating minimization algorithm for which stronger convergence results can be established, under wide assumptions. In recent years, proximal methods have become increasingly popular for solving inverse problems in image processing [6] due to their ability to tackle minimization problems involving sums of possibly nonsmooth functions, such as those arising in the presence of hard or sparsity promoting constraints. However, most of the existing work on proximal methods has focused on data recovery problems based on a model involving a linear operator assumed to be known a priori.

Throughout the paper, we use the following notation. $\Gamma_0(\mathcal{H})$ denotes the class of lower semicontinuous convex functions from a real Hilbert space $\mathcal{H}$ to $]-\infty,+\infty]$. An example of a function in $\Gamma_0(\mathcal{H})$ is the indicator function ι_C of a nonempty closed convex subset C of $\mathcal{H}$, which takes on the value 0 on C and $+\infty$ on $\mathcal{H}\smallsetminus C$. A fundamental tool for the derivation of the algorithms in this paper is the proximity operator of a function $f\in\Gamma_0(\mathcal{H})$, which is defined as

$$\operatorname{prox}_f\colon\mathcal{H}\to\mathcal{H}\colon x\mapsto\operatorname*{argmin}_{y\in\mathcal{H}} f(y)+\frac{1}{2}\|x-y\|^2. \quad (1)$$

For background on proximity operators and their use in signal and image processing problems, the reader is referred to [6, 7]. Subsequently, two Hilbert spaces will be of interest: the standard Euclidean space $\mathbb{R}^N$ and $\mathbb{R}^{M\times N}$, the space of real valued matrices of size $M\times N$ endowed with the Fröbenius norm. For notational conciseness, the norms of both spaces will be denoted by $\|\cdot\|$.

In Section 2, the blind data recovery problem under consideration is formulated. In Section 3, we emphasize some of the limitations of basic alternating minimization schemes. The new proximal optimization method is introduced in Section 4. Finally, in Section 5, we apply the proposed algorithm to a blind image deconvolution problem.

2. PROBLEM

We consider the standard linear observation model

$$z=\overline{L}\overline{x}+w, \quad (2)$$

where $z\in\mathbb{R}^M$ is the observed data, $\overline{L}\in\mathbb{R}^{M\times N}$ models to the linear measurement process, $\overline{x}\in\mathbb{R}^N$ is the target data and $w\in\mathbb{R}^N$ is some noise perturbation. Our objective is to recover $\overline{x}$ from z, without knowledge of $\overline{L}$. Such a problem arises in many blind data recovery problems in deconvolution, source separation or reconstruction. An estimate of $(\overline{L},\overline{x})$ is obtained by solving the following optimization problem.

Problem 2.1 Set $\Phi\colon(x,L)\mapsto f(x)+g(L)+h(z-Lx)$, where $h\colon\mathbb{R}^M\to\mathbb{R}$ is a differentiable convex function which has a Lipschitz continuous gradient over every bounded subsets of $\mathbb{R}^M$, $f\in\Gamma_0(\mathbb{R}^N)$, and $g\in\Gamma_0(\mathbb{R}^{M\times N})$. The objective is to

$$\operatorname*{minimize}_{x\in\mathbb{R}^N,\ L\in\mathbb{R}^{M\times N}}\ \Phi(x,L). \quad (3)$$

This work was supported by the Agence Nationale de la Recherche under grants ANR-08-BLAN-0294-02 and ANR-09-EMER-004-03.

For example, in a Bayesian framework, a solution to Problem 3 is a Maximum A Posteriori estimate of $(\overline{L}, \overline{x})$ if one assumes that w is a realization of a random vector with probability density function $\propto \exp(-h(\cdot))$, $\overline{L}$ is a realization of a random matrix with probability density function $\propto \exp(-g(\cdot))$, $\overline{x}$ is a realization of a random vector with probability density $\propto \exp(-f(\cdot))$, and the three latter random variables are jointly independent.

Often, f and g can be decomposed as sums of simpler functions, say $f = \sum_{i=1}^{p} f_i$ and $g = \sum_{j=1}^{q} g_j$, where $(f_i)_{1\leq i\leq p}$ are functions in $\Gamma_0(\mathbb{R}^N)$, and $(g_j)_{1\leq j\leq q}$ are functions in $\Gamma_0(\mathbb{R}^{M\times N})$. Problem (3) then becomes

$$\underset{x\in\mathbb{R}^N,\, L\in\mathbb{R}^{M\times N}}{\text{minimize}} \quad \sum_{i=1}^{p} f_i(x) + \sum_{j=1}^{q} g_j(L) + h(z - Lx). \quad (4)$$

Because of the coupling term $(x, L) \mapsto h(z - Lx)$, the objective function is in general not convex. In the supervised case when $\overline{L}$ (respectively, $\overline{x}$) is known a natural choice is to set $q = 1$ and $g_1 = \iota_{\{0\}}(\cdot - \overline{L})$ (respectively, $p = 1$ and $f_1 = \iota_{\{0\}}(\cdot - \overline{x})$). In such instances, (4) reduces to a classical convex problem.

3. LIMITATIONS OF BASIC ALTERNATING MINIMIZATION PROCEDURES

Let us define the following auxiliary functions. For every $L \in \mathbb{R}^{M\times N}$, we set

$$\varphi_L \colon x \mapsto \sum_{i=1}^{p} f_i(x) + h(z - Lx) \quad (5)$$

and, for every $x \in \mathbb{R}^N$, we set

$$\psi_x \colon L \mapsto \sum_{j=1}^{q} g_j(L) + h(z - Lx), \quad (6)$$

For fixed values of L and x, we have $\varphi_L \in \Gamma_0(\mathbb{R}^N)$ and $\psi_x \in \Gamma_0(\mathbb{R}^{M\times N})$.

A popular approach for solving Problem 3 consists of applying an alternating minimization approach. The corresponding algorithm, sometimes called the Gauss-Seidel method, takes the following form.

Algorithm 3.1

Fix $L_0 \in \mathbb{R}^{M\times N}$
For $k = 0, 1, \ldots$
$\quad x_k \in \operatorname{Argmin} \varphi_{L_k}$
$\quad L_{k+1} \in \operatorname{Argmin} \psi_{x_k}$.

This algorithm may provide satisfactory results in practice. However, it is well known that such an alternating minimization procedure requires quite restrictive conditions to guarantee convergence to a local minimizer, e.g., [4] (the lack of convergence of alternating minimization procedures can also be observed with convex objectives). In the present context, a simple counterexample is the following.

Example 3.2 Assume that $N = M$ and set

$$f = \iota_C,\ g = \|\cdot\|_1 + \iota_D, \quad \text{and} \quad h = \frac{1}{2}\|\cdot\|^2, \quad (7)$$

where $C = [-1, 1]^N$ and where D is the vector subspace of diagonal matrices of $\mathbb{R}^{N\times N}$. If we suppose that $z \in C$ and initialize Algorithm 3.1 with $L_0 = 0$, a resulting sequence of iterates is given by

$$(\forall k \in \mathbb{N}) \qquad x_k = (-1)^k [1, \ldots, 1]^\top,\ L_k = 0. \quad (8)$$

Hence, $(x_k)_{k\in\mathbb{N}}$ does not converge.

4. PROPOSED OPTIMIZATION METHOD

As an alternative to Algorithm 3.1, we propose to use the following alternating proximal algorithm generating a sequence $(x_k, L_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^N \times \mathbb{R}^{M\times N}$:

Algorithm 4.1

Fix $x_0 \in \mathbb{R}^N, L_0 \in \mathbb{R}^{M\times N}$ and $]\underline{\rho}, \overline{\rho}[\subset]0, +\infty[$
For $k = 0, 1, \ldots$
$\quad (\lambda_k, \mu_k) \in [\underline{\rho}, \overline{\rho}]^2$
$\quad x_{k+1} = \operatorname{prox}_{\lambda_k \varphi_{L_k}} x_k$
$\quad L_{k+1} = \operatorname{prox}_{\mu_k \psi_{x_{k+1}}} L_k$.

It is worth pointing out that, in the supervised case when $q = 1$ and $g_1 = \iota_{\{0\}}(\cdot - \overline{L})$ (respectively, $p = 1$ and $f_1 = \iota_{\{0\}}(\cdot - \overline{x})$), the method reduces to the standard proximal point algorithm [6].

The computational complexity of Algorithm 4.1 is usually similar to that of Algorithm 3.1. In addition, Algorithm 4.1 enjoys attractive convergence properties. First, we recall that Φ is coercive if $\lim_{\|x\|+\|L\|\to+\infty} \Phi(x, L) = +\infty$ and it is semi-algebraic if its graph $\operatorname{gra} \Phi = \big\{\big((x, L), \nu\big) \,\big|\, \nu = \Phi(x, L)\big\}$ is a semi-algebraic set, that is, it can be expressed as a finite union of subsets of $(\mathbb{R}^N \times \mathbb{R}^{M\times N}) \times \mathbb{R}$ defined by a finite number of polynomial inequalities. The set of semi-algebraic functions constitutes a wide class of functions, including many standard functions, and it is stable through common operations (e.g., addition, multiplication, inversion, and composition). The following result follows from Lemma 5 and Theorem 9 in [2].

Proposition 4.2 *Let Φ be as in Problem 2.1. Then, for every $k \in \mathbb{N}$,*

$$\Phi(x_{k+1}, L_{k+1}) + \frac{1}{2\lambda_k}\|x_{k+1} - x_k\|^2 + \frac{1}{2\mu_k}\|L_{k+1} - L_k\|^2 \leq \Phi(x_k, L_k). \quad (9)$$

If, in addition, Φ is coercive, then Φ has a global minimizer. If, furthermore, Φ is semi-algebraic, then every sequence $(x_k, L_k)_{k\in\mathbb{N}}$ generated by Algorithm 4.1 converges to a critical point of Φ.

Proposition 4.3 [2, Theorem 11] *Let Φ be as in Problem 2.1 and suppose that it is coercive and semi-algebraic. Let $(\widetilde{x}, \widetilde{L})$ be the limit of a sequence $(x_k, L_k)_{k\in\mathbb{N}}$ generated by Algorithm 4.1. Then one of the following holds.*

(i) *Convergence occurs in a finite number of iterations.*

(ii) *There exist $\tau \in]0,1[$ and $\eta \in]0,+\infty[$ such that, for every $k \in \mathbb{N}$, $\|x_k - \widetilde{x}\|^2 + \|L_k - \widetilde{L}\|^2 \leq \eta\tau^k$.*

(iii) *There exist θ and η in $]0,+\infty[$ such that, for every $k \in \mathbb{N} \smallsetminus \{0\}$, $\|x_k - \widetilde{x}\|^2 + \|L_k - \widetilde{L}\|^2 \leq \eta k^{-\theta}$.*

A main difficulty in the implementation of Algorithm 4.1 is the computation of the proximity operators $\operatorname{prox}_{\lambda_k \varphi_{L_k}}$ and $\operatorname{prox}_{\mu_k \psi_{x_{k+1}}}$ at each iteration k. This task can be efficiently performed by using the parallel Dykstra-like proximal algorithm proposed in [5]. This leads to the following routine to compute $\operatorname{prox}_{\lambda\varphi_L} x$ with $\lambda \in]0,+\infty[$, $L \in \mathbb{R}^{M\times N}$, and $x \in \mathbb{R}^N$ (a similar method can be employed to compute $\operatorname{prox}_{\mu\psi_x} L$ with $\mu \in]0,+\infty[$, $x \in \mathbb{R}^N$, and $L \in \mathbb{R}^{M\times N}$).

Algorithm 4.4

Fix $y_0 = x, s_{1,0} = y_0, \ldots, s_{p+1,0} = y_0,$ and
$(\omega_i)_{1\leq i\leq p+1} \in]0,1]^{p+1}$ such that $\sum_{i=1}^{p+1} \omega_i = 1$
For $\ell = 0, 1, \ldots$
  For $i = 1, \ldots, p$
    $r_{i,\ell} = \operatorname{prox}_{\frac{\lambda f_i}{\omega_i}} s_{i,\ell}$
  $r_{p+1,\ell} = \operatorname{prox}_{\frac{\lambda h(z-L\cdot)}{\omega_{p+1}}} s_{p+1,\ell}$
  $y_{\ell+1} = \sum_{i=1}^{p+1} \omega_i r_{i,\ell}$
  For $i = 1, \ldots, p+1$
    $s_{i,\ell+1} = y_{\ell+1} + s_{i,\ell} - r_{i,\ell}$.

Proposition 4.5 *[5, Theorem 4.2] The sequence $(y_\ell)_{\ell\in\mathbb{N}}$ generated by Algorithm 4.4 converges to $\operatorname{prox}_{\lambda\varphi_L} x$.*

5. SIMULATION EXAMPLE

We consider a blind deconvolution scenario where an original 8 bit $N_1 \times N_2$ image $\overline{x}$ is degraded by a blur and the addition of a zero-mean white Gaussian noise with variance σ^2. We have thus $M = N = N_1 N_2$ and $h = \|\cdot\|^2/(2\sigma^2)$. A classical generalized Gaussian frame-analysis prior [1, 8] is assumed for the original image, which yields

$$(\forall x \in \mathbb{R}^N) \quad f_1(x) = \sum_{\ell=1}^{K} \zeta_\ell |(Fx)^{(\ell)}|^{\kappa_\ell}, \tag{10}$$

where $F \in \mathbb{R}^{K\times N}$ corresponds to a frame analysis operator, $(\zeta_\ell)_{1\leq\ell\leq K} \in [0,+\infty[^K$ and $(\kappa_\ell)_{1\leq\ell\leq K} \in [1,+\infty[^K$ (rational values of $(\kappa_\ell)_{1\leq\ell\leq K}$ are chosen so that f_1 is a semi-algebraic function). We also take into account the available information on the range intensity values by setting $f_2 = \iota_{[0,255]^N}$. Hence, $p = 2$.

The blur is modeled by a periodic convolution with a kernel $\overline{H} \in \mathbb{R}^{P\times Q}$. Let S be the linear operator which maps a filter kernel $H \in \mathbb{R}^{P\times Q}$ to its associated circulant block-circulant transform matrix of size $N \times N$ (when $N_1 \geq P$ and $N_2 \geq Q$). This yields $\overline{L} = S(\overline{H})$. Prior information on the unknown degradation operator can be incorporated by assuming that $L = S(H)$, where $H = (H_{n,m})_{1\leq n\leq P, 1\leq m\leq Q}$ satisfies the following properties.

- nonnegativity: $H \in ([0,+\infty[)^{P\times Q}$.
- mean: $\sum_{n=1}^{P}\sum_{m=1}^{Q} H_{n,m} = 1$.
- bounds on vertical variations of the blur:

$$(\forall n \in \{1,\ldots,P-1\})(\forall m \in \{1,\ldots,Q\}) \quad \alpha_{1,n,m} \leq H_{n+1,m} - H_{n,m} \leq \beta_{1,n,m}, \tag{11}$$

  where $(\alpha_{1,n,m})_{n,m}$ and $(\beta_{1,n,m})_{n,m}$ are given.

- bounds on horizontal variations of the blur:

$$(\forall n \in \{1,\ldots,P\})(\forall m \in \{1,\ldots,Q-1\}) \quad \alpha_{2,n,m} \leq H_{n,m+1} - H_{n,m} \leq \beta_{2,n,m}, \tag{12}$$

  where $(\alpha_{2,n,m})_{n,m}$ and $(\beta_{2,n,m})_{n,m}$ are given.

The above constraints define four closed convex subsets $(D_j)_{1\leq j\leq 4}$ of $\mathbb{R}^{P\times Q}$. We consequently choose $q = 4$ and $(\forall j \in \{1,2,3,4\})\, g_j = \iota_{S(D_j)}$.

Fig. 1 displays the original satellite image (with $N_1 = N_2 = 512$) which is blurred by an anisotropic truncated Gaussian kernel of size 7×7. The blurred signal-to-noise ratio is equal to 20.7 dB in the degraded image shown in Fig 2. Fig. 3 shows the result provided by Algorithm 4.1. The method was initialized with the blurred image ($x_0 = z$) and a uniform kernel. A symlet 8 wavelet basis decomposition computed over 4 resolution levels is used in this example, and the parameters $(\zeta_\ell)_{1\leq\ell\leq K}$ and $(\kappa_\ell)_{1\leq\ell\leq K}$ are subband-dependent (they have been chosen with a maximum likelihood approach). The bounds on the vertical (respectively, horizontal) variations of the blur are $(\forall m \in \{1,\ldots,7\})$ $(\forall n \in \{1,\ldots,3\})$ $\alpha_{1,n,m} = 0$ and $\beta_{1,n,m} = 6 \times 10^{-3}$ and $(\forall n \in \{4,\ldots,6\})$ $\alpha_{1,n,m} = -6 \times 10^{-3}$ and $\beta_{1,n,m} = 0$ (respectively, $(\forall n \in \{1,\ldots,7\})$ $(\forall m \in \{1,\ldots,3\})$ $\alpha_{2,n,m} = 0$ and $\beta_{2,n,m} = 3 \times 10^{-3}$ and $(\forall m \in \{4,\ldots,6\})$ $\alpha_{2,n,m} = -3 \times 10^{-3}$ and $\beta_{2,n,m} = 0$). As shown visually, and confirmed by the provided signal-to-noise ratios (SNR), the results are close to those obtained by a similar wavelet-based restoration approach which assumes that the blur is known (see Fig. 4).

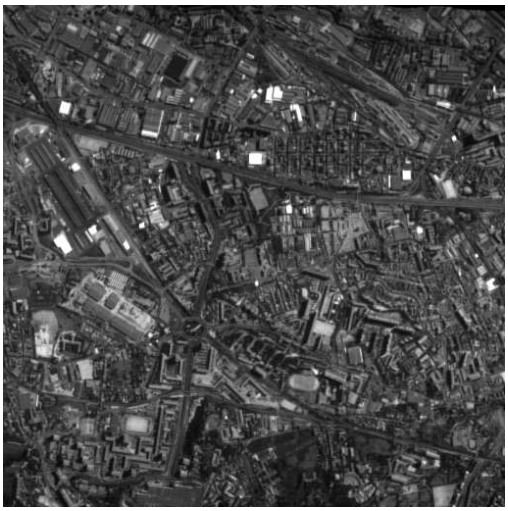

Fig. 1. Original image $\overline{x}$.

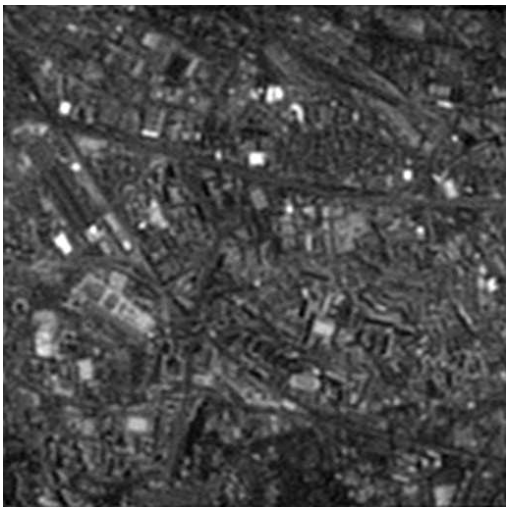

Fig. 3. Blind restoration: SNR = 14.4 dB, SSIM = 0.807.

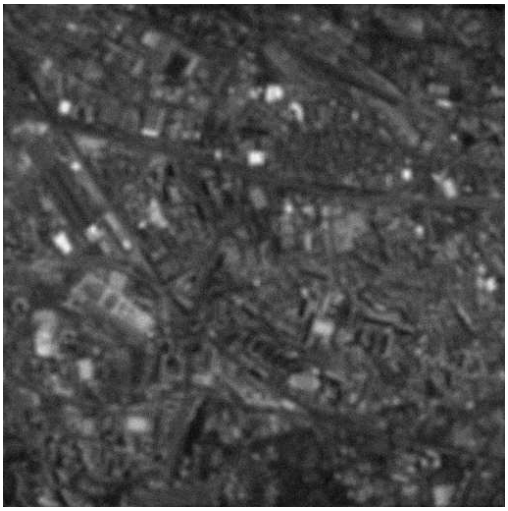

Fig. 2. Degraded image z: SNR = 12.5 dB, SSIM = 0.683.

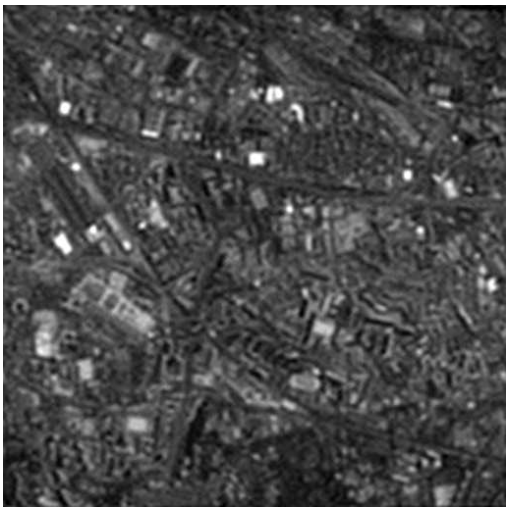

Fig. 4. Supervised restoration: SNR = 14.5 dB, SSIM = 0.812.

6. REFERENCES

[1] A. Antoniadis, D. Leporini, and J.-C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statist. Neerlandica*, vol. 56, pp. 434–453, 2002.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, pp. 438–457, 2010.

[3] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Process.*, vol. 18, pp. 12–26, 2009.

[4] T. F. Chan and C. K. Wong, "Convergence of the alternating minimization algorithm for blind deconvolution," *Linear Algebra Appl.*, vol. 316, pp. 259–285, 2000.

[5] P. L. Combettes, "Iterative construction of the resolvent of a sum of maximal monotone operators," *J. Convex Anal.*, vol. 16, pp. 727–748, 2009.

[6] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, (H. H. Bauschke et al., Eds.). New York: Springer-Verlag, 2010.

[7] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.

[8] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947–968, 2007.

[9] E. Y. Lam and J. W. Goodman, "Iterative statistical approach to blind image deconvolution," *J. Opt. Soc. Amer. A*, vol. 17, pp. 1177–1184, 2000.