

PAPER • OPEN ACCESS

A Bregman inexact linesearch-based forward-backward algorithm for nonsmooth nonconvex optimization

To cite this article: Simone Rebegoldi *et al* 2018 *J. Phys.: Conf. Ser.* **1131** 012013

View the [article online](#) for updates and enhancements.

You may also like

- [Convex optimization algorithms in medical image reconstruction—in the age of AI](#)
Jingyan Xu and Frédéric Noo
- [Inexact Bregman iteration with an application to Poisson data reconstruction](#)
A Benfenati and V Ruggiero
- [Variational regularisation for inverse problems with imperfect forward operators and general noise models](#)
Leon Bungert, Martin Burger, Yury Korolev et al.

PRIME
PACIFIC RIM MEETING
ON ELECTROCHEMICAL
AND SOLID STATE SCIENCE

HONOLULU, HI
Oct 6–11, 2024

Abstract submission deadline:
April 12, 2024

Learn more and submit!

Joint Meeting of
The Electrochemical Society
•
The Electrochemical Society of Japan
•
Korea Electrochemical Society

A Bregman inexact linesearch-based forward-backward algorithm for nonsmooth nonconvex optimization

Simone Rebegoldi¹, Silvia Bonettini² and Marco Prato²

¹ Dipartimento di Matematica e Informatica, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy

² Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/b, 41125 Modena, Italy

E-mail: simone.rebegoldi@unife.it

Abstract. In this paper, we present a forward-backward linesearch-based algorithm suited for the minimization of the sum of a smooth (possibly nonconvex) function and a convex (possibly nonsmooth) term. Such algorithm first computes inexactly the proximal operator with respect to a given Bregman distance, and then ensures a sufficient decrease condition by performing a linesearch along the descent direction. The proposed approach can be seen as an instance of the more general class of descent methods presented in [7], however, unlike in [7], we do not assume the strong convexity of the Bregman distance used in the proximal evaluation. We prove that each limit point of the iterates sequence is stationary, we show how to compute an approximate proximal-gradient point with respect to a Bregman distance and, finally, we report the good numerical performance of the algorithm on a large scale image restoration problem.

1. Introduction

This paper is concerned with the following optimization problem

$$\arg \min_{x \in \mathbb{R}^n} f(x) \equiv f_0(x) + f_1(x) \quad (1)$$

where $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable on an open set Ω , $f_1 : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is a proper, convex, continuous function on a closed domain $\text{dom } f_1 \subseteq \Omega$, and f is bounded from below. Several problems arising in signal and image processing can be reformulated as (1), such as image deconvolution, image segmentation, image compression and inpainting, non-negative matrix and tensor factorization [4, 10]. The structure of the objective function in (1) can be exploited by using the forward-backward (or proximal-gradient) algorithm [7, 12, 13], whose general iteration is given by

$$x^{(k+1)} = x^{(k)} + \lambda_k \left(\text{prox}_{\alpha_k f_1}^{D_k} (x^{(k)} - \alpha_k D_k^{-1} \nabla f_0(x^{(k)})) - x^{(k)} \right), \quad (2)$$

where $\alpha_k > 0$ is a steplength parameter, $D_k \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, $\text{prox}_{\alpha f_1}^D$ denotes the proximal operator of αf_1 w.r.t. the norm $\|x\|_D = x^T D x$ induced by D , i.e.

$$\text{prox}_{\alpha f_1}^D(z) = \arg \min_{y \in \mathbb{R}^n} f_1(y) + \frac{1}{2\alpha} \|y - z\|_D^2, \quad (3)$$



and $\lambda_k \in (0, 1)$ is an additional relaxation parameter which aims at ensuring the sufficient decrease of the objective function. Since, in its basic implementation, method (2) can be quite slow, several accelerated versions have been devised, in which either the steplength α_k and/or the scaling matrix D_k are chosen in a variable manner [7, 8, 11]. More recently, the scheme (2) has been generalized by replacing the (scaled) Euclidean norm in (3) with the more general concept of Bregman distance [1, 7, 9]. Given $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ a strictly convex and differentiable function on $\text{int}(\text{dom } \varphi) \neq \emptyset$, the Bregman distance $D_\varphi : \text{dom}(\varphi) \times \text{int}(\text{dom } \varphi) \rightarrow [0, +\infty]$ associated to φ is defined as [2, Definition 3.1]

$$D_\varphi(x, y) = \varphi(x) - \varphi(y) - \nabla\varphi(y)^T(x - y), \quad \forall x \in \text{dom}(\varphi), \quad \forall y \in \text{int}(\text{dom } \varphi). \quad (4)$$

Observe that the scaled Euclidean norm is recovered from (4) by setting $\varphi = \|\cdot\|_D^2/2$. Then one can apply (2) by using, instead of (3), the following generalized proximal operator [1, p. 337]

$$\text{prox}_{\alpha f_1}^\varphi(z) = \arg \min_{y \in \mathbb{R}^n} f_1(y) + \frac{1}{\alpha} D_\varphi(y, z). \quad (5)$$

In the following, we present a Bregman version of algorithm (2) whose main features are the computation of an approximate proximal–gradient point $\tilde{y}^{(k)}$ with respect to a prefixed Bregman distance, the adaptive choice of the steplength α_k , which can take any value in a prefixed interval $[\alpha_{\min}, \alpha_{\max}]$, and the linesearch performed along the descent direction detected by the approximate point $\tilde{y}^{(k)}$. Such algorithm can be considered as an instance of the general framework proposed in [7], where the concept of Bregman distance is replaced by a generalized distance-like function satisfying appropriate assumptions. We prove that each limit point of the iterates sequence is stationary under the assumption of super-coercivity of the Bregman distance, which is weaker than the one of strong convexity adopted in [7]. Furthermore, we provide a general procedure to compute an approximate Bregman proximal–gradient point satisfying the required inexactness criterion, and we evaluate the effectiveness of the proposed approach on a Poisson image restoration problem.

2. Algorithm and convergence result

The domain of a function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is defined as $\text{dom } f = \{x \in \mathbb{R}^n : f(x) < \infty\}$. Given $\epsilon > 0$, the ϵ -subdifferential of a proper, convex function f at $x \in \text{dom } f$ is the set $\partial_\epsilon f(x) = \{w \in \mathbb{R}^n : f(z) \geq f(x) + w^T(z - x) - \epsilon, \forall z \in \mathbb{R}^n\}$. The limiting subdifferential of a function f at $x \in \text{dom } f$ is defined as $\partial f(x) = \{v \in \mathbb{R}^n : \exists \{y^{(k)}\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n, v^{(k)} \in \hat{\partial} f(y^{(k)}) \forall k \in \mathbb{N} \text{ s.t. } y^{(k)} \rightarrow x, f(y^{(k)}) \rightarrow f(x) \text{ and } v^{(k)} \rightarrow v\}$, where $\hat{\partial} f(x)$ denotes the Fréchet subdifferential of f at x [8, Definition 1]. The conjugate function f^* is defined as $f^*(v) = \sup_x x^T v - f(x)$. For any $c \in \mathbb{R}$, $\text{lev}_{\leq c}(f) = \{x \in \mathbb{R}^n : f(x) \leq c\}$ denotes the sublevel set of f .

Throughout the paper, we will make the following blanket assumptions.

Assumption 1 (i) φ is a Legendre function [2], i.e. it is strictly convex and differentiable on $\text{int}(\text{dom } \varphi)$, and $\|\nabla\varphi(x^{(k)})\| \rightarrow \infty$ for every sequence $\{x^{(k)}\} \subseteq \text{int}(\text{dom } \varphi)$ converging to a boundary point of $\text{dom } \varphi$ as $k \rightarrow \infty$.

(ii) $0 \in \text{int}(\text{dom } \varphi)$.

(iii) $\text{dom } f_1 \subseteq \text{int}(\text{dom } \varphi)$.

(iv) $\text{dom } \varphi^* = \mathbb{R}^n$.

Remark 1 Note that Assumption 1(iv) is equivalent to require that φ is super-coercive, namely [2, Definition 2.15, Proposition 2.16]

$$\lim_{\|x\| \rightarrow +\infty} \frac{\varphi(x)}{\|x\|} = +\infty. \quad (6)$$

Such an hypothesis is weaker than requiring strong convexity, i.e. that $\varphi - \frac{m}{2}\|\cdot\|^2$ is convex for a certain $m > 0$. Indeed every strongly convex function is super-coercive [3, Corollary 11.17] but the converse is not true. As an example, consider the Boltzmann–Shannon entropy $\varphi(x) = \sum_{i=1}^n \varphi_i(x_i) = \sum_{i=1}^n (x_i + \gamma) \log(x_i + \gamma)$, where $\gamma > 0$ and $0 \log 0 = 0$ [1]. Note that $\varphi^*(v) = \sum_{i=1}^n (e^{v_i-1} - \gamma v_i)$ is defined on all \mathbb{R}^n and, therefore, φ is super-coercive. However, the second derivative of φ_i is bounded away from zero only on bounded subsets of the half-line $(-\gamma, +\infty)$, hence φ is not strongly convex.

We also recall the following properties of Bregman distances.

Theorem 1 [2] Suppose that $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ satisfies Assumption 1. Then

- (i) $\nabla\varphi : \text{int}(\text{dom } \varphi) \rightarrow \text{int}(\text{dom } \varphi^*)$ is a bijection and $(\nabla\varphi)^{-1} = \nabla\varphi^*$.
- (ii) For all $x, y \in \text{int}(\text{dom } \varphi)$, we have

$$D_\varphi(x, y) = D_{\varphi^*}(\nabla\varphi(y), \nabla\varphi(x)). \quad (7)$$

- (iii) (Three Points Identity) For all $x \in \text{dom}(\varphi)$, $y, z \in \text{int}(\text{dom } \varphi)$, we have

$$D_\varphi(x, z) = D_\varphi(x, y) + D_\varphi(y, z) + (\nabla\varphi(y) - \nabla\varphi(z))^T(x - y). \quad (8)$$

Given $x^{(k)} \in \text{dom } f_1$ and $\alpha_k > 0$, we define the following function

$$h_{\alpha_k}(y, x^{(k)}) := \nabla f_0(x^{(k)})^T(y - x^{(k)}) + \frac{1}{\alpha_k} D_\varphi(y, x^{(k)}) + f_1(y) - f_1(x^{(k)}) \quad (9)$$

and, assuming that (9) has an unique minimizer, we set $y^{(k)} = \arg \min_{y \in \mathbb{R}^n} h_{\alpha_k}(y, x^{(k)})$. As described in [1, Section 3.1], the point $y^{(k)}$ can be naturally decomposed as a Bregman proximal–gradient point. Indeed, let us write down the optimality condition

$$0 \in \partial h_{\alpha_k}(y^{(k)}, x^{(k)}) \Leftrightarrow 0 \in \alpha_k \nabla f_0(x^{(k)}) + \nabla\varphi(y^{(k)}) - \nabla\varphi(x^{(k)}) + \alpha_k \partial f_1(y^{(k)}). \quad (10)$$

Since Assumption 1(iv) implies that $\nabla\varphi(x^{(k)}) - \alpha_k \nabla f_0(x^{(k)}) \in \text{dom } \nabla\varphi^*$, we can always define the point $z^{(k)} = \nabla\varphi^*(\nabla\varphi(x^{(k)}) - \alpha_k \nabla f_0(x^{(k)}))$. Then, by recalling Theorem 1(i), the optimality condition (10) can be equivalently rewritten as

$$0 \in \nabla\varphi(y^{(k)}) - \nabla\varphi(z^{(k)}) + \alpha_k \partial f_1(y^{(k)})$$

which yields $y^{(k)} = \text{prox}_{\alpha_k f_1}^\varphi(z^{(k)})$.

We are now ready to fully detail the proposed approach, which is reported in Algorithm 1. At each iteration, we look for an approximation $\tilde{y}^{(k)}$ of the Bregman proximal–gradient point $y^{(k)}$ which satisfies the following conditions

$$h_{\alpha_k}(\tilde{y}^{(k)}, x^{(k)}) < 0, \quad 0 \in \partial_{\epsilon_k} h_{\alpha_k}(\tilde{y}^{(k)}, x^{(k)}). \quad (11)$$

On one hand, the first condition guarantees that the point $\tilde{y}^{(k)}$ defines a descent direction $d^{(k)} = \tilde{y}^{(k)} - x^{(k)}$ [7, Proposition 2.2.] and, on the other hand, it ensures that the linesearch at STEP 3 is well-defined [7, Proposition 3.1]. The second condition can be seen as a relaxation of the optimality condition (10) associated to the point $y^{(k)}$, where the exact subdifferential of h_{α_k} is replaced with its ϵ_k –subdifferential. The linesearch at STEP 3 enforces condition (12), thus making $\{f(x^{(k)})\}_{k \in \mathbb{N}}$ a monotone nondecreasing sequence.

Algorithm 1 belongs to the class of line–search based descent methods described in [7], denominated Variable Metric Inexact Linesearch based Algorithms (VMILA). In [7, Theorem

Algorithm 1 Bregman Inexact Linesearch based Algorithm

Choose $x^{(0)} \in \text{dom } f_1$, $0 \leq \alpha_{\min} \leq \alpha_{\max}$, $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ satisfying Assumption 1, $\delta, \beta \in (0, 1)$.
 FOR $k = 0, 1, 2, \dots$

STEP 1. Choose $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and $\epsilon_k > 0$.

STEP 2. Compute $\tilde{y}^{(k)}$ satisfying (11) and set $d^{(k)} = \tilde{y}^{(k)} - x^{(k)}$.

STEP 3. Compute the smallest non-negative integer i_k such that

$$f(x^{(k)} + \delta^{i_k} d^{(k)}) \leq f(x^{(k)}) + \beta \delta^{i_k} h_{\alpha_k}(\tilde{y}^{(k)}, x^{(k)}) \quad (12)$$

and set $\lambda_k = \delta^{i_k}$.

STEP 4. Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$.

END

3.1], the stationarity of the limit points of VMILA is proved according to the following outline: (i) show that the subsequence $\{\tilde{y}^{(k)}\}_{k \in K}$ is bounded (using the strong convexity of φ); (ii) deduce that $\lim_{k \rightarrow \infty, k \in K} h_{\alpha_k}(\tilde{y}^{(k)}, x^{(k)}) = 0$ (using [7, Proposition 3.1]); (iii) conclude that the limit point is stationary (using again the strong convexity of φ). In the following theorem, we prove an analogous result for Algorithm 1, by adopting the same line of proof described above, but assuming that φ is super-coercive instead of strongly convex.

Theorem 2 Assume that there exists a limit point \bar{x} of $\{x^{(k)}\}_{k \in \mathbb{N}}$ and that $\lim_{k \rightarrow \infty} \epsilon_k = 0$. Then \bar{x} is a stationary point for problem (1), i.e. $0 \in \partial f(\bar{x})$.

Proof: Let $K \subseteq \mathbb{N}$ be a subset of indices such that $\lim_{k \rightarrow \infty, k \in K} x^{(k)} = \bar{x}$ and $\lim_{k \rightarrow \infty, k \in K} \alpha_k = \bar{\alpha} \in [\alpha_{\min}, \alpha_{\max}]$. By continuity of the function $h_{\alpha}(y, x)$ with respect to the arguments α and x , we also have $\lim_{k \rightarrow \infty, k \in K} y^{(k)} = \bar{y}$, with $\bar{y} = \arg \min_y h_{\bar{\alpha}}(y, \bar{x})$. From (10), it follows $w \in \partial f_1(y^{(k)})$ if and only if $w = -\nabla f_0(x^{(k)}) - \frac{1}{\alpha_k}(\nabla \varphi(y^{(k)}) - \nabla \varphi(x^{(k)}))$. Observe that the right-hand side condition in (11) can be rewritten as $h_{\alpha_k}(\tilde{y}^{(k)}) - h_{\alpha_k}(y^{(k)}) \leq \epsilon_k$. Then we have

$$\begin{aligned} \epsilon_k &\geq h_{\alpha_k}(\tilde{y}^{(k)}) - h_{\alpha_k}(y^{(k)}) & (13) \\ &= \nabla f_0(x^{(k)})^T (\tilde{y}^{(k)} - y^{(k)}) + \frac{1}{\alpha_k} \left(D_{\varphi}(\tilde{y}^{(k)}, x^{(k)}) - D_{\varphi}(y^{(k)}, x^{(k)}) \right) + f_1(\tilde{y}^{(k)}) - f_1(y^{(k)}) \\ &\geq \nabla f_0(x^{(k)})^T (\tilde{y}^{(k)} - y^{(k)}) + \frac{1}{\alpha_k} \left(D_{\varphi}(\tilde{y}^{(k)}, x^{(k)}) - D_{\varphi}(y^{(k)}, x^{(k)}) \right) + w^T (\tilde{y}^{(k)} - y^{(k)}) \\ &= \frac{1}{\alpha_k} \left(D_{\varphi}(\tilde{y}^{(k)}, x^{(k)}) - D_{\varphi}(y^{(k)}, x^{(k)}) - (\nabla \varphi(y^{(k)}) - \nabla \varphi(x^{(k)}))^T (\tilde{y}^{(k)} - y^{(k)}) \right) \\ &\stackrel{(8)}{=} \frac{1}{\alpha_k} D_{\varphi}(\tilde{y}^{(k)}, y^{(k)}) \geq \frac{1}{\alpha_{\max}} D_{\varphi}(\tilde{y}^{(k)}, y^{(k)}). \end{aligned}$$

Thanks to the hypothesis on $\{\epsilon_k\}_{k \in \mathbb{N}}$, the above inequalities yield $\lim_{k \rightarrow \infty} D_{\varphi}(\tilde{y}^{(k)}, y^{(k)}) = 0$. By applying (8) with $x = \tilde{y}^{(k)}$, $y = \bar{y}$, $z = y^{(k)}$, and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} D_{\varphi}(\tilde{y}^{(k)}, y^{(k)}) &= D_{\varphi}(\tilde{y}^{(k)}, \bar{y}) + D_{\varphi}(\bar{y}, y^{(k)}) + (\nabla \varphi(\bar{y}) - \nabla \varphi(y^{(k)}))^T (\tilde{y}^{(k)} - \bar{y}) \\ &\geq D_{\varphi}(\tilde{y}^{(k)}, \bar{y}) - \|\nabla \varphi(\bar{y}) - \nabla \varphi(y^{(k)})\| \|\tilde{y}^{(k)} - \bar{y}\| \end{aligned}$$

Since $\{y^{(k)}\}_{k \in K}$ is bounded, $\nabla \varphi$ is continuous and $\lim_{k \rightarrow \infty, k \in K} D_{\varphi}(\tilde{y}^{(k)}, \bar{y}) = 0$, there exist two constants $c_1, c_2 > 0$ such that

$$\frac{D_{\varphi}(\tilde{y}^{(k)}, \bar{y}) - c_1}{\|\tilde{y}^{(k)}\|} \leq c_2, \quad \forall k \in K$$

or, in other words, $\{\tilde{y}^{(k)}\}_{k \in K} \subseteq \text{lev}_{\leq c_2} \left(\frac{D_\varphi(\cdot, \bar{y}) - c_1}{\|\cdot\|} \right)$. By Assumption 1(iv) combined with Remark 1, this means that the sequence $\{\tilde{y}^{(k)}\}_{k \in K}$ lies in a sublevel set of a coercive function. Therefore, $\{\tilde{y}^{(k)}\}_{k \in K}$ is bounded and, consequently, we can apply [7, Proposition 3.1] to obtain that $\lim_{k \rightarrow \infty, k \in K} h_{\alpha_k}(\tilde{y}^{(k)}, x^{(k)}) = 0$. This last fact, combined with (13) and the assumption on $\{\epsilon_k\}_{k \in \mathbb{N}}$, implies $\lim_{k \rightarrow \infty, k \in K} h_{\alpha_k}(y^{(k)}, x^{(k)}) = h_{\bar{\alpha}}(\bar{y}, \bar{x}) = 0$, which allows to conclude that \bar{x} is stationary [7, Proposition 2.3]. \square

3. Practical computation of the approximate point $\tilde{y}^{(k)}$

In this section, we show how we can practically compute a point $\tilde{y}^{(k)}$ satisfying (11) in the case when $f_1 = g \circ A$, being $A \in \mathbb{R}^{m \times n}$ and $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\infty\}$ a proper, convex function. The following procedure is a generalization of the one presented in [7, Section 4.2] for (scaled) Euclidean distances. Let us rewrite (10) in its equivalent primal-dual and dual formulations:

$$\min_{y \in \mathbb{R}^n} h_{\alpha_k}(y, x^{(k)}) = \min_{y \in \mathbb{R}^n} \max_{v \in \mathbb{R}^m} F_{\alpha_k}(y, v, x^{(k)}) = \max_{v \in \mathbb{R}^m} \Psi_{\alpha_k}(v, x^{(k)}).$$

By applying Theorem 1(i) to point $z^{(k)}$, we have $\nabla f_0(x^{(k)}) = (\nabla \varphi(x^{(k)}) - \nabla \varphi(z^{(k)})) / \alpha_k$. Using this relation and (8), we can rewrite the function $h_{\alpha_k}(\cdot, x^{(k)})$ as follows

$$h_{\alpha_k}(y, x^{(k)}) = \frac{1}{\alpha_k} \left(D_\varphi(y, z^{(k)}) - D_\varphi(x^{(k)}, z^{(k)}) \right) + f_1(y) - f_1(x^{(k)}).$$

Recalling that $f_1(y) = \max_{v \in \mathbb{R}^m} v^T A y - g^*(v)$, we come to the expression of $F_{\alpha_k}(y, v, x^{(k)})$:

$$F_{\alpha_k}(y, v, x^{(k)}) = \frac{1}{\alpha_k} D_\varphi(y, z^{(k)}) + v^T A y - g^*(v) - \frac{1}{\alpha_k} D_\varphi(x^{(k)}, z^{(k)}) - f_1(x^{(k)}). \quad (14)$$

The minimum of $F_{\alpha_k}(\cdot, v, x^{(k)})$ is reached at the point $y_{\min}^{(k)} = \nabla \varphi^*(\nabla \varphi(z^{(k)}) - \alpha_k A^T v)$. Replacing $y_{\min}^{(k)}$ into $F_{\alpha_k}(y, v, x^{(k)})$, we get the dual function $\Psi_{\alpha_k}(\cdot, x^{(k)})$:

$$\begin{aligned} \Psi_{\alpha_k}(v, x^{(k)}) &= \frac{1}{\alpha_k} D_\varphi(y_{\min}^{(k)}, z^{(k)}) + v^T A y_{\min}^{(k)} - g^*(v) - \frac{1}{\alpha_k} D_\varphi(x^{(k)}, z^{(k)}) - f_1(x^{(k)}) \\ &\stackrel{(8)}{=} \frac{1}{\alpha_k} D_\varphi(\nabla \varphi^*(\nabla \varphi(z^{(k)}) - \alpha_k A^T v), 0) + \frac{1}{\alpha_k} (\nabla \varphi(0) - (\nabla \varphi(z^{(k)}) - \alpha_k A^T v))^T (\nabla \varphi^*(\nabla \varphi(z^{(k)}) - \alpha_k A^T v)) \\ &\quad - g^*(v) - \frac{1}{\alpha_k} D_\varphi(x^{(k)}, z^{(k)}) + \frac{1}{\alpha_k} D_\varphi(0, z^{(k)}) - f_1(x^{(k)}) \\ &\stackrel{(7)}{=} \frac{1}{\alpha_k} D_{\varphi^*}(\nabla \varphi(0), \nabla \varphi(z^{(k)}) - \alpha_k A^T v) \\ &\quad + \frac{1}{\alpha_k} (\nabla \varphi(0) - (\nabla \varphi(z^{(k)}) - \alpha_k A^T v))^T (\nabla \varphi^*(\nabla \varphi(z^{(k)}) - \alpha_k A^T v) - \nabla \varphi^*(\nabla \varphi(0))) \\ &\quad - g^*(v) - \frac{1}{\alpha_k} D_\varphi(x^{(k)}, z^{(k)}) + \frac{1}{\alpha_k} D_\varphi(0, z^{(k)}) - f_1(x^{(k)}) \\ &\stackrel{(8)}{=} -\frac{1}{\alpha_k} D_{\varphi^*}(\nabla \varphi(z^{(k)}) - \alpha_k A^T v, \nabla \varphi(0)) - g^*(v) - \frac{1}{\alpha_k} D_\varphi(x^{(k)}, z^{(k)}) + \frac{1}{\alpha_k} D_\varphi(0, z^{(k)}) - f_1(x^{(k)}) \end{aligned}$$

where, in the second equality, we used (8) with $x = y_{\min}^{(k)}$, $y = 0$, $z = z^{(k)}$ whereas, in the fourth equality, we used (8) with $x = z = \nabla \varphi(0)$, $y = \nabla \varphi(z^{(k)}) - \alpha_k A^T v$. If it is possible to generate a sequence $\{v^{(k, \ell)}\}_{\ell \in \mathbb{N}}$ such that $v^{(k, \ell)} \xrightarrow{\ell \rightarrow \infty} \arg \max_v \Psi_{\alpha_k}(v, x^{(k)})$,

$\Psi_{\alpha_k}(v^{(k,\ell)}, x^{(k)}) \xrightarrow{\ell \rightarrow \infty} \max_v \Psi_{\alpha_k}(v, x^{(k)})$, then a point $\tilde{y}^{(k)}$ satisfying (11) can be found by setting $\tilde{y}^{(k,\ell)} = P_{\text{dom } f_1}(\nabla \varphi^*(\nabla \varphi(z^{(k)}) - \alpha_k A^T v^{(k,\ell)})$ and stopping the iterates when condition

$$h_{\alpha_k}(\tilde{y}^{(k,\ell)}, x^{(k)}) \leq \eta \Psi_{\alpha_k}(v^{(k,\ell)}, x^{(k)}) \quad (15)$$

is satisfied, with $\eta \in (0, 1]$ being a prefixed parameter. Indeed, $\tilde{y}^{(k)} = \tilde{y}^{(k,\ell)}$ satisfies (11) with $\epsilon_k = (1 - 1/\eta)h_{\alpha_k}(\tilde{y}^{(k)}, x^{(k)})$ [7, Section 4.5].

4. Application to Poisson image restoration

In the Bayesian framework, the recovering of an unknown image $x \in \mathbb{R}^n$ from a blurred noisy image $g \in \mathbb{R}^n$ can be reformulated as a problem of the form (1), where f_0 is a discrepancy function, typically depending on the type of noise affecting the data, whereas f_1 is a regularization term including a-priori information and possible physical constraints. When the data are corrupted by Poisson noise, f_0 is the generalized Kullback-Leibler (KL) divergence [4]:

$$f_0(x) = KL(Hx + b; g) = \sum_{i=1}^n g_i \log \frac{g_i}{(Hx)_i + b} + (Hx)_i + b - g_i \quad (16)$$

where the matrix $H \in \mathbb{R}^{n \times n}$ represents the blurring operator satisfying the conditions $H_{i,j} \geq 0$, $H\mathbf{1}$ and $H^T\mathbf{1}$ (with $\mathbf{1}$ being the vector of all ones) and $b > 0$ is a background term. In order to preserve the edges of the image and the non negativity of its pixels, the term f_1 is chosen as

$$f_1(x) = \rho TV(x) + \iota_{\mathbb{R}_{\geq 0}^n}(x), \quad (17)$$

where $TV(x) = \rho \sum_{i=1}^n \|\nabla_i x\|_2$ is the total variation functional [14], being $\nabla_i \in \mathbb{R}^{2 \times n}$ the discrete gradient operator at the i -th pixel, $\iota_{\mathbb{R}_{\geq 0}^n}$ is the indicator function of the non negative orthant and $\rho > 0$ is the regularization parameter.

We now discuss the application of Algorithm 1 to the above problem. Concerning the Legendre function φ , it seems quite natural, looking at the discrepancy term (16), to use the Boltzmann–Shannon entropy described in Remark 1. Indeed, the Bregman distance associated to such function is $D_\varphi(x, y) = \sum_{i=1}^n (x_i + \gamma) \log \left(\frac{x_i + \gamma}{y_i + \gamma} \right) + (y_i + \gamma) - (x_i + \gamma) = KL(y + \gamma\mathbf{1}; x + \gamma\mathbf{1})$, i.e. the Kullback–Leibler divergence itself. The approximate point $\tilde{y}^{(k)}$ is then computed as described in Section 3, using the scaled gradient projection method [6] to generate the dual sequence $\{v^{(k,\ell)}\}_{\ell \in \mathbb{N}}$. Regarding the steplength α_k , we observe that the point $z^{(k)}$ can be rewritten as $z_i^{(k)} = x_i^{(k)} - D_k(\alpha_k) \nabla_i f_0(x^{(k)}) = x_i^{(k)} - q_i(\alpha_k)(x_i + \gamma) \nabla_i f_0(x^{(k)})$, with $q_i(\alpha) = (1 - e^{-\alpha \nabla_i f_0(x^{(k)})}) / \nabla_i f_0(x^{(k)})$ when $\nabla_i f_0(x^{(k)}) \neq 0$ and $q_i(\alpha) = \alpha$ when $\nabla_i f_0(x^{(k)}) = 0$. Thus, it is reasonable to consider the following quasi-Newton approaches

$$\alpha_1^{(k)} = \min_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} \|D_k(\alpha)^{-1} s^{(k)} - w^{(k)}\|^2, \quad \alpha_2^{(k)} = \min_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} \|s^{(k)} - D_k(\alpha) w^{(k)}\|^2$$

where $s^{(k)} = x^{(k)} - x^{(k-1)}$ and $w^{(k)} = \nabla f_0(x^{(k)}) - \nabla f_0(x^{(k-1)})$. In our experiments, we alternate the two above rules using the same strategy adopted in [6].

We show the results obtained by running Algorithm 1 on a set of three test problems well-known in the literature, whose description can be found in [7]. In Figure 1 we report the relative decrease of the objective function with respect to the iteration number of Algorithm 1, denominated VMILA-Bregman, in comparison with two other instances of the VMILA algorithm, one in which the proximal operator (5) is computed w.r.t to the Euclidean norm (VMILA-ID), and the other one in which a variable scaled Euclidean norm is used (VMILA-SG). As it can be drawn from the plots, both VMILA-Bregman and VMILA-SG outperform VMILA-ID of several orders of magnitude, which suggests that adopting non Euclidean metrics may be advantageous both in terms of efficiency and accuracy.

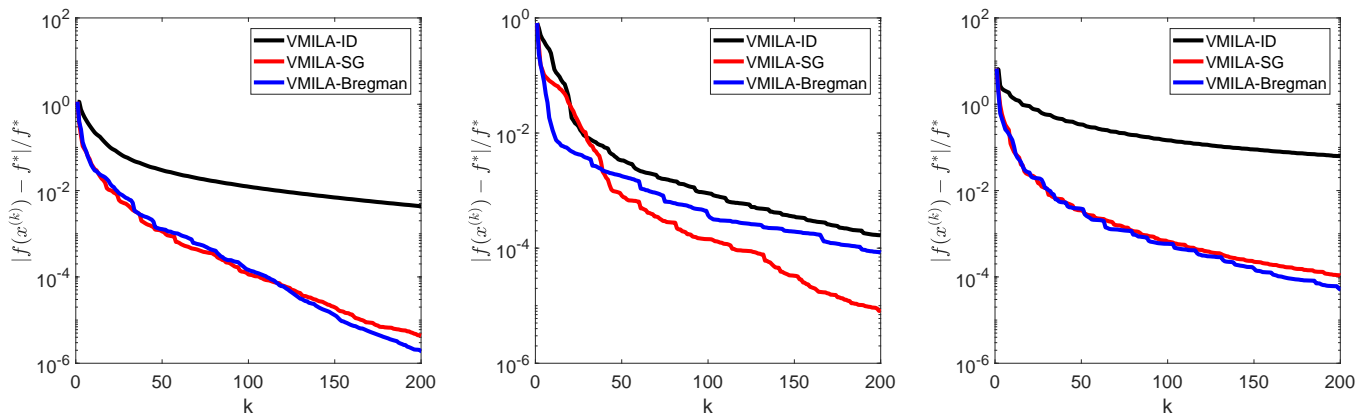


Figure 1. Relative decrease of the objective function values with respect to the iteration number. Left column: cameraman. Middle column. micro. Right column: phantom.

5. Acknowledgements

The authors are members of the INdAM Research group GNCS. This work has been partially supported by the Italian GNCS-INdAM under the project GNCS - Finanziamento Giovani Ricercatori 2017-2018.

References

- [1] H. H. Bauschke, J. Bolte, and M. Teboulle 2017, A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications, *Math. Oper. Res.* **4**(1), 330–348.
- [2] H. H. Bauschke, and J. M. Borwein 1997, Legendre functions and the method of Bregman projection, *J. Convex Anal.* **4**, 27–67.
- [3] H. H. Bauschke, and P. L. Combettes 2011, *Convex analysis and monotone operator theory in Hilbert spaces*, (Springer).
- [4] M. Bertero, and P. Boccacci 1998, *Introduction to Inverse Problems in Imaging*, (Bristol: Institute of Physics Publishing).
- [5] M. Bertero, P. Boccacci, G. Desider, and G. Vicidomini 2009, Image deblurring with Poisson data: From cells to galaxies, *Inverse Probl.* **25**(12), 123006.
- [6] S. Bonettini, R. Zanella, and L. Zanni 2009, A scaled gradient projection method for constrained image deblurring, *Inverse Probl.* **25**, 015002.
- [7] S. Bonettini, I. Loris, F. Porta, and M. Prato 2016, Variable metric inexact line-search-based methods for nonsmooth optimization, *SIAM J. Optim.* **26**(2), 891–921.
- [8] S. Bonettini, I. Loris, F. Porta, M. Prato, and S. Rebegoldi 2017, On the convergence of a linesearch based proximal–gradient method for nonconvex optimization, *Inverse Probl.* **33**(5), 055005.
- [9] S. Bonettini, M. Prato, and S. Rebegoldi 2016, A cyclic block coordinate descent method with generalized gradient projections, *Appl. Math. Comput.* **286**, 288–300.
- [10] A. Chambolle, and T. Pock 2016, An introduction to continuous optimization for imaging, *Acta Numerica* **25**, 161–319.
- [11] E. Chouzenoux, J.-C. Pesquet, and A. Repetti 2014, Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function, *J. Optim. Theory Appl.* **162**(1), 107–132.
- [12] P. L. Combettes, and V. R. Wajs 2005, Signal recovery by proximal forward–backward splitting, *Multiscale Model. Simul.* **4**, 1168–1200.
- [13] P. L. Combettes, and J.-C. Pesquet 2011, Proximal splitting methods in signal processing, in *Fixed-point algorithms for inverse problems in science and engineering*, H. H. Bauschke et al., New York, Springer, 185–212.
- [14] L. Rudin, S. Osher, and E. Fatemi 1992, Nonlinear total variation based noise removal algorithms, *J. Phys. D.* **60**(1-4), 259–268.