

This is the peer reviewed version of the following article:

A convergent least-squares regularized blind deconvolution approach / Cornelio, Anastasia; Porta, Federica; Prato, Marco. - In: APPLIED MATHEMATICS AND COMPUTATION. - ISSN 0096-3003. - STAMPA. - 259:(2015), pp. 173-186. [10.1016/j.amc.2015.02.048]

*Terms of use:*

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

12/01/2026 19:09

# A convergent least-squares regularized blind deconvolution approach

Anastasia Cornelio<sup>a</sup>, Federica Porta<sup>b</sup>, Marco Prato<sup>b,\*</sup>

<sup>a</sup>*Dipartimento di Scienze e Metodi dell'Ingegneria, Università di Modena e Reggio Emilia, Via Amendola 2 Pad. Morselli, 42122 Reggio Emilia, Italy*

<sup>b</sup>*Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/b, 41125 Modena, Italy*

---

## Abstract

The aim of this work is to present a new and efficient optimization method for the solution of blind deconvolution problems with data corrupted by Gaussian noise, which can be reformulated as a constrained minimization problem whose unknowns are the point spread function (PSF) of the acquisition system and the true image. The objective function we consider is the weighted sum of the least-squares fit-to-data discrepancy and possible regularization terms accounting for specific features to be preserved in both the image and the PSF. The solution of the corresponding minimization problem is addressed by means of a proximal alternating linearized minimization (PALM) algorithm, in which the updating procedure is made up of one step of a gradient projection method along the arc and the choice of the parameter identifying the steplength in the descent direction is performed automatically by exploiting the optimality conditions of the problem. The resulting approach is a particular case of a general scheme whose convergence to stationary points of the constrained minimization problem has been recently proved. The effectiveness of the iterative method is validated in several numerical simulations in image reconstruction problems.

**Keywords:** Blind deconvolution, Regularization, Least squares problems, Optimization methods

**2010 MSC:** 65F22, 65K05, 65R32, 94A08

---

## 1. Introduction

Image deconvolution is an extremely prolific field which on one hand finds applications in a large variety of areas (physics, medicine, engineering,...) and on the other hand rounds up the efforts of a large community of mathematicians working on inverse problems and optimization methods. Most of the resulting works deal with the ill-conditioned discrete problem in which the blurring matrix is assumed to be known and the goal is to find a good approximation of the unknown image by means of some regularization approaches [1]. However, in many real applications the blurring matrix is not completely known due to a lack of information on the acquisition model and/or to external agents which corrupt the measured image (atmospheric turbulence, thermal blooming,...). This situation is known as *blind deconvolution* and most strategies to approach this problem are based on a simultaneous recovery of both the image approximation and the point spread function (PSF) of the acquisition system. Blind deconvolution is a very actual field and a much more challenging problem than the image deconvolution one, due to the strongly ill-posedness caused by the non-uniqueness of the solution. A review of some recent results can be found e.g. in two recent papers by Almeida & Figueiredo and Oliveira et al. [2, 3], even if many other different approaches have been developed. If the noise corrupting the measured data is assumed to have a Gaussian nature, blind deconvolution can be addressed by means of the constrained minimization of the squared Euclidean norm of the residuals plus some regularization terms suitably chosen according to the object and PSF to be reconstructed. An alternative formulation involves the unconstrained minimization of the same objective function with the addition of the indicator functions of the feasible sets, which can be numerically solved with forward-backward

---

\*Corresponding author. Tel: +39 059 2055590, Fax: + 39 059 2055216

Email addresses: anastasia.cornelio@unimore.it (Anastasia Cornelio), federica.porta@unimore.it (Federica Porta), marco.prato@unimore.it (Marco Prato)

Preprint submitted to Applied Mathematics and Computation

July 25, 2017

splitting methods [4, 5].

The starting point of this paper is a proximal alternating method recently proposed by Bolte et al. [6] for a more general class of nonconvex and nonsmooth minimization problems, in which the parameters defining the method are fixed by using the Lipschitz constants of the partial gradients of the least-squares plus regularization part of the objective function. Convergence of the resulting sequence to a stationary point is ensured by the Kurdyka–Łojasiewicz property [7, 8, 9]. In particular, for the specific case of blind deconvolution, we extend the convergence results proved in [6] to a wider range of parameters, which allows the corresponding scheme to converge much faster toward the limit point. Moreover, we introduce a practical adaptive choice of the parameters based on a measure of the optimality condition violation, and we test the proposed algorithm in some simulated numerical experiments.

The paper is organized as follows: in Section 2 we introduce the blind deconvolution from Gaussian data and we recall the specific formulation of the proximal alternating linearized minimization proposed in [6] for this problem. Our proposed extension of the scheme is described in Section 3, together with the analogous convergence results and some hints for the choice of the parameters. Section 4 is devoted to some numerical experiments on synthetic datasets, while our conclusions are given in Section 5.

## 2. Problem setting

When dealing with Gaussian noise, blind deconvolution can be modeled as the minimization problem

$$\min_{\mathbf{x} \in X, \mathbf{h} \in H} F(\mathbf{x}, \mathbf{h}) = \frac{1}{2} \|\mathbf{h} * \mathbf{x} - \mathbf{g}\|^2 + \lambda_1 R_1(\mathbf{x}) + \lambda_2 R_2(\mathbf{h}), \quad (1)$$

where  $X \subseteq \mathbb{R}^n$  and  $H \subseteq \mathbb{R}^m$  are the nonempty, closed and convex feasible sets of the unknown image  $\mathbf{x}$  and PSF  $\mathbf{h}$ , respectively,  $*$  denotes the convolution operator,  $\mathbf{g}$  is the measured image,  $R_1, R_2$  are differentiable regularization terms and  $\lambda_1, \lambda_2$  are positive regularization parameters. As usual, the computation of the convolution product is performed by means of a matrix–vector multiplication  $\mathbf{h} * \mathbf{x} = \mathbf{H}\mathbf{x} = \mathbf{X}\mathbf{h}$ , where  $\mathbf{X}$  and  $\mathbf{H}$  are suitable structured matrices depending on the choice of the boundary conditions [10]. Examples of regularization terms frequently used in the applications are the following:

- $R^{\text{T0}}(\mathbf{z}) = \|\mathbf{z}\|^2$  (Tikhonov regularization of order 0);
- $R^{\text{T1}}(\mathbf{z}) = \|\nabla \mathbf{z}\|^2$  (Tikhonov regularization of order 1);
- $R^{\text{HS}}(\mathbf{z}) = \sum_i \sqrt{\|(\nabla \mathbf{z})_i\|^2 + \beta^2}$  (hypersurface regularization), where  $(\nabla \mathbf{z})_i$  is the 2-vector representing the discretization of the gradient of  $\mathbf{z}$  at pixel  $i$  in the horizontal and vertical directions.

As concerns the feasible sets, we consider non-negative images and non-negative and normalized PSFs, therefore we assume

$$\begin{aligned} X &= \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x} \geq \mathbf{0}\}, \\ H &= \{\mathbf{h} \in \mathbb{R}^m | \mathbf{h} \geq \mathbf{0}, \sum_i h_i = 1\}. \end{aligned}$$

If we denote with  $\mathcal{I}_X$  and  $\mathcal{I}_H$  the indicator functions of  $X$  and  $H$ , respectively, then problem (1) can be rewritten as

$$\min_{\mathbf{x} \in \mathbb{R}^n, \mathbf{h} \in \mathbb{R}^m} \Psi(\mathbf{x}, \mathbf{h}) = F(\mathbf{x}, \mathbf{h}) + \mathcal{I}_X(\mathbf{x}) + \mathcal{I}_H(\mathbf{h}). \quad (2)$$

Problem (2) can be addressed by means of recently proposed optimization methods, provided that the function  $F$  and its gradient satisfy some properties that we recall in the following definition.

**Definition 1** We define the set  $\mathcal{F}$  of functions  $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  such that:

- (i)  $\inf_{\mathbf{x} \in \mathbb{R}^n, \mathbf{h} \in \mathbb{R}^m} F(\mathbf{x}, \mathbf{h}) > -\infty$ ;

(ii) for any  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{h} \in \mathbb{R}^m$ , the partial gradients  $\nabla_{\mathbf{x}}F(\cdot, \mathbf{h})$  and  $\nabla_{\mathbf{h}}F(\mathbf{x}, \cdot)$  are globally Lipschitz continuous, i.e., there exist  $L_{\mathbf{x}}(\mathbf{h}), L_{\mathbf{h}}(\mathbf{x}) > 0$  such that

$$\begin{aligned}\|\nabla_{\mathbf{x}}F(\mathbf{x}_1, \mathbf{h}) - \nabla_{\mathbf{x}}F(\mathbf{x}_2, \mathbf{h})\| &\leq L_{\mathbf{x}}(\mathbf{h})\|\mathbf{x}_1 - \mathbf{x}_2\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n, \\ \|\nabla_{\mathbf{h}}F(\mathbf{x}, \mathbf{h}_1) - \nabla_{\mathbf{h}}F(\mathbf{x}, \mathbf{h}_2)\| &\leq L_{\mathbf{h}}(\mathbf{x})\|\mathbf{h}_1 - \mathbf{h}_2\| \quad \forall \mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^m.\end{aligned}$$

Moreover, for each bounded subset  $B_1 \times B_2$  of  $\mathbb{R}^n \times \mathbb{R}^m$ , there exist  $\alpha(B_1), \alpha(B_2) > 0$  such that

$$\begin{aligned}\sup\{L_{\mathbf{h}}(\mathbf{x}) : \mathbf{x} \in B_1\} &\leq \alpha(B_1), \\ \sup\{L_{\mathbf{x}}(\mathbf{h}) : \mathbf{h} \in B_2\} &\leq \alpha(B_2);\end{aligned}$$

(iii) the gradient  $\nabla F$  is locally Lipschitz continuous, i.e., for each bounded subset  $B$  of  $\mathbb{R}^n \times \mathbb{R}^m$ , there exists  $M(B) > 0$  such that

$$\|\nabla F(\mathbf{x}_1, \mathbf{h}_1) - \nabla F(\mathbf{x}_2, \mathbf{h}_2)\| \leq M(B)\|(\mathbf{x}_1 - \mathbf{x}_2, \mathbf{h}_1 - \mathbf{h}_2)\|$$

for all  $(\mathbf{x}_1, \mathbf{h}_1), (\mathbf{x}_2, \mathbf{h}_2) \in B$ .

The function  $F$  defined in problem (1), when  $R_1, R_2$  are the Tikhonov or HS regularizations, belongs to the set  $\mathcal{F}$ . Indeed, in this case  $F$  is readily lower bounded. Moreover, for Tikhonov regularization of order 0 we have

$$\begin{aligned}\|\nabla_{\mathbf{x}}F(\mathbf{x}_1, \mathbf{h}) - \nabla_{\mathbf{x}}F(\mathbf{x}_2, \mathbf{h})\| &= \|(\mathbf{H}^T \mathbf{H} + 2\lambda_1 \mathbf{I})(\mathbf{x}_1 - \mathbf{x}_2)\| \\ &\leq \|\mathbf{H}^T \mathbf{H} + 2\lambda_1 \mathbf{I}\| \|\mathbf{x}_1 - \mathbf{x}_2\|,\end{aligned}$$

thus giving  $L_{\mathbf{x}}(\mathbf{h}) = \|\mathbf{H}\|^2 + 2\lambda_1$  (and in a similar way  $L_{\mathbf{h}}(\mathbf{x}) = \|\mathbf{X}\|^2 + 2\lambda_2$ ). As for Tikhonov regularization of order 1 and HS, we start by defining the  $2d \times d$  matrix (where  $d = n$  for the image and  $d = m$  for the PSF)

$$\mathbf{D} = [\mathbf{D}_1^T \cdots \mathbf{D}_d^T]^T,$$

where  $\mathbf{D}_i$  is the  $2 \times d$  matrix defined such that  $\mathbf{D}_i \mathbf{z}$  is the forward difference approximation of the gradient at the  $i$ -th component of a vector  $\mathbf{z} \in \mathbb{R}^d$ . Each row of  $\mathbf{D}_i$  ( $i = 1, \dots, d$ ) contains two nonzero elements and it results that  $\|\mathbf{D}\|^2 \leq 8$  [11]. The regularization operators and their gradients can be rewritten by means of  $\mathbf{D}$ :

$$\begin{aligned}R^{\text{T1}}(\mathbf{x}) &= \|\mathbf{D}\mathbf{x}\|^2 \quad ; \quad \nabla_{\mathbf{x}}R^{\text{T1}}(\mathbf{x}) = 2\mathbf{D}^T \mathbf{D}\mathbf{x}, \\ R^{\text{HS}}(\mathbf{x}) &= \sum_i \sqrt{\|\mathbf{D}_i \mathbf{x}\|^2 + \beta^2} \quad ; \quad \nabla_{\mathbf{x}}R^{\text{HS}}(\mathbf{x}) = \sum_i \frac{\mathbf{D}_i^T \mathbf{D}_i \mathbf{x}}{\sqrt{\|\mathbf{D}_i \mathbf{x}\|^2 + \beta^2}}.\end{aligned}$$

It follows that for Tikhonov regularization of order 1 we have

$$L_{\mathbf{x}}(\mathbf{h}) \leq \|\mathbf{H}\|^2 + 16\lambda_1 \quad ; \quad L_{\mathbf{h}}(\mathbf{x}) \leq \|\mathbf{X}\|^2 + 16\lambda_2,$$

while for HS regularization [12]

$$\begin{aligned}L_{\mathbf{x}}(\mathbf{h}) &= \|\mathbf{H}\|^2 + \lambda_1 \|\mathbf{D}\|^2 / \beta^2 \leq \|\mathbf{H}\|^2 + 8\lambda_1 / \beta^2, \\ L_{\mathbf{h}}(\mathbf{x}) &= \|\mathbf{X}\|^2 + \lambda_2 \|\mathbf{D}\|^2 / \beta^2 \leq \|\mathbf{X}\|^2 + 8\lambda_2 / \beta^2.\end{aligned}$$

Finally, the last part of condition (ii) and condition (iii) straightly follow from the fact that  $F$  is a  $C^2$  function (see Remark 3(iv) of [6]). Under these considerations, the proximal version

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \arg \min_{\mathbf{x} \in \mathbb{R}^n} \langle \mathbf{x} - \mathbf{x}^{(k)}, \nabla_{\mathbf{x}}F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) \rangle + \frac{c_k}{2} \|\mathbf{x} - \mathbf{x}^{(k)}\|^2 + \mathcal{I}_X(\mathbf{x}), \\ \mathbf{h}^{(k+1)} &= \arg \min_{\mathbf{h} \in \mathbb{R}^m} \langle \mathbf{h} - \mathbf{h}^{(k)}, \nabla_{\mathbf{h}}F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) \rangle + \frac{d_k}{2} \|\mathbf{h} - \mathbf{h}^{(k)}\|^2 + \mathcal{I}_H(\mathbf{h}),\end{aligned} \tag{3}$$

being  $c_k, d_k$  positive real numbers, of the usual Gauss–Seidel scheme can be exploited for the solution of (2) [13, 5, 14]. In particular, in a recent paper [6] the authors proposed a proximal alternating linearized minimization (PALM) algorithm which, starting from a given initial point  $(\mathbf{x}^{(0)}, \mathbf{h}^{(0)}) \in X \times H$ , generates a sequence  $(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})$  ( $k = 0, 1, \dots$ ) in the following way:

- (i) take  $\gamma_x > 1$  and set  $c_k = \gamma_x L_x(\mathbf{h}^{(k)})$ ;
- (ii) choose  $\mathbf{x}^{(k+1)} \in \text{prox}_{c_k}^{\mathcal{I}_X} \left( \mathbf{x}^{(k)} - \frac{1}{c_k} \nabla_x F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) \right)$ ;
- (iii) take  $\gamma_h > 1$  and set  $d_k = \gamma_h L_h(\mathbf{x}^{(k+1)})$ ;
- (iv) choose  $\mathbf{h}^{(k+1)} \in \text{prox}_{d_k}^{\mathcal{I}_H} \left( \mathbf{h}^{(k)} - \frac{1}{d_k} \nabla_h F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) \right)$ .

We recall that the proximal operator exploited in (ii) and (iv) is, in general, defined as

$$\text{prox}_{\theta}^{\sigma}(\mathbf{z}) = \arg \min_{\mathbf{u} \in \mathbb{R}^d} \sigma(\mathbf{u}) + \frac{\theta}{2} \|\mathbf{u} - \mathbf{z}\|^2$$

for a given proper and lower semicontinuous function  $\sigma : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ ,  $\theta > 0$  and  $\mathbf{z} \in \mathbb{R}^d$ . Since in our case  $X, H$  are convex, the proximal operators associated to  $\mathcal{I}_X$  and  $\mathcal{I}_H$  reduce to the projections  $\mathcal{P}_X$  and  $\mathcal{P}_H$  on the sets  $X$  and  $H$ , respectively. The inverse of the steplength parameters  $c_k, d_k$  must be chosen in order to satisfy the sufficient decrease properties [6]

$$F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) \leq F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) + \langle \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \nabla_x F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) \rangle + \frac{c_k}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \quad (4)$$

$$F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)}) \leq F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) + \langle \mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}, \nabla_h F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) \rangle + \frac{d_k}{2} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|^2 \quad (5)$$

which, in particular, can be surely reached by choosing  $c_k \geq L_x(\mathbf{h}^{(k)})$ ,  $d_k \geq L_h(\mathbf{x}^{(k+1)})$  [13]. These conditions combined with the requirement that  $F$  satisfies the Kurdyka–Łojasiewicz (KL) property [7, 8, 9] allow to demonstrate that the sequence generated by the PALM method converges to a critical point.

**Definition 2** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function. For  $-\infty < v_1 < v_2 \leq +\infty$ , let us set

$$[v_1 < f < v_2] = \{\mathbf{z} \in \mathbb{R}^d : v_1 < f(\mathbf{z}) < v_2\}.$$

Moreover, we denote with  $\partial f(\mathbf{z})$  the subdifferential of  $f$  at  $\mathbf{z} \in \mathbb{R}^d$  (see [15, Definition 2.1]) and with  $\text{dist}(\mathbf{z}, \Omega)$  the distance between a point  $\mathbf{z}$  and a set  $\Omega \subset \mathbb{R}^d$ . The function  $f$  is said to have the KL property at  $\bar{\mathbf{z}} \in \text{dom} \partial f := \{\mathbf{z} \in \mathbb{R}^d : \partial f(\mathbf{z}) \neq \emptyset\}$  if there exist  $v \in (0, +\infty]$ , a neighborhood  $U$  of  $\bar{\mathbf{z}}$  and a continuous concave function  $\phi : [0, v) \rightarrow (0, +\infty)$  such that:

- $\phi(0) = 0$ ;
- $\phi$  is  $C^1$  on  $(0, v)$ ;
- for all  $s \in (0, v)$ ,  $\phi'(s) > 0$ ;
- for all  $\mathbf{z} \in U \cap [f(\bar{\mathbf{z}}) < f < f(\bar{\mathbf{z}}) + v]$ , the KL inequality holds:

$$\phi'(f(\mathbf{z}) - f(\bar{\mathbf{z}})) \text{dist}(\mathbf{z}, \partial f(\bar{\mathbf{z}})) \geq 1.$$

If  $f$  satisfies the KL property at each point of  $\text{dom} \partial f$  then  $f$  is called a KL function.

The KL condition concerns a large variety of functions which we do not consider in our work [15, 6, 16]. We only recall that any  $p$ -norm is a KL function (see e.g. [6, Section 5] and [16, Section 2.2]), which includes all the possible objective functions we consider in our blind deconvolution framework. Unfortunately, parameters  $c_k, d_k$  greater than or equal to the Lipschitz constants often lead to very small steps, with a result of a very slow convergence of the related sequence toward the critical point itself. In the next sections we extend the convergence result of the PALM method to parameters  $c_k, d_k$  which satisfy (4) and (5) but are smaller than the Lipschitz constants of the partial gradients (thus allowing larger steplengths) and we propose a linesearch strategy for their automatic selection.

### 3. The modified PALM approach

#### 3.1. Parameters selection

The rule we choose to update the constants  $c_k, d_k$  is the one used in [17] in the context of non-negative matrix factorization and is based on the optimality conditions for each subproblem. More in details, if we define the projected gradient of the function  $F$  as

$$\nabla^P F(\mathbf{x}, \mathbf{h}) = (\nabla_x^P F(\mathbf{x}, \mathbf{h}), \nabla_h^P F(\mathbf{x}, \mathbf{h})),$$

where

$$\begin{aligned}\nabla_x^P F(\mathbf{x}, \mathbf{h}) &= \mathcal{P}_X(\mathbf{x} - \nabla_x F(\mathbf{x}, \mathbf{h})) - \mathbf{x}, \\ \nabla_h^P F(\mathbf{x}, \mathbf{h}) &= \mathcal{P}_H(\mathbf{h} - \nabla_h F(\mathbf{x}, \mathbf{h})) - \mathbf{h},\end{aligned}$$

then a point  $(\mathbf{x}^*, \mathbf{h}^*)$  is stationary for problem (2) if and only if we have

$$\nabla^P F(\mathbf{x}^*, \mathbf{h}^*) = \mathbf{0}. \quad (6)$$

Thus, the quantities  $\|\nabla_x^P F(\mathbf{x}, \mathbf{h})\|, \|\nabla_h^P F(\mathbf{x}, \mathbf{h})\|$  can be considered as a measure of the optimality of the point  $(\mathbf{x}, \mathbf{h})$  and exploited to estimate the steplength parameters at each iteration. Inspired by [17], we start from  $c_k = L_x(\mathbf{h}^{(k)})$ ,  $d_k = L_h(\mathbf{x}^{(k+1)})$  and we multiply them for a factor  $\delta < 1$  until conditions (4) and (5) are violated or

$$\|\nabla_x^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)})\| \leq \eta_x^{(k)}, \quad \|\nabla_h^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})\| \leq \eta_h^{(k+1)}, \quad (7)$$

where the adaptive tolerances  $\eta_x^{(k)}, \eta_h^{(k+1)}$  are initialized as

$$\eta_x^{(0)} = \eta_h^{(0)} = 10^{-3} \cdot \|\nabla^P F(\mathbf{x}^{(0)}, \mathbf{h}^{(0)})\|$$

and are updated as

$$\eta_x^{(k+1)} = \begin{cases} 0.1 \cdot \eta_x^{(k)} & \text{if } \eta_x^{(k)} \geq \|\nabla_x^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})\| \\ \eta_x^{(k)} & \text{otherwise} \end{cases}, \quad (8)$$

$$\eta_h^{(k+1)} = \begin{cases} 0.1 \cdot \eta_h^{(k)} & \text{if } \eta_h^{(k)} \geq \|\nabla_h^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)})\| \\ \eta_h^{(k)} & \text{otherwise} \end{cases}. \quad (9)$$

The idea of this first step is that the Lipschitz constants often provide parameters which largely satisfy conditions (4) and (5), and therefore these parameters are reduced by looking at the optimality conditions until a sort of neighborhood of the “optimal” values is found. If the reduction step leads to a violation of (4) and (5) before satisfying (7), the parameters  $c_k, d_k$  are slightly increased by a factor  $\mu > 1$  until the former inequalities return to hold true.

Algorithm 1 gives the scheme of the resulting modified PALM method, which reduces to the standard PALM scheme if one neglects Steps 3, 4, 7 and 8. A graphical scheme of the proposed rule to select the  $c_k, d_k$  parameters is shown in Figure 1. The iterative procedure is stopped when one of the following conditions has been satisfied:

- objective function:  $|F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)}) - F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)})| < \varepsilon_1 |F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})|$ ;
- norm of the projected gradient:  $\|\nabla^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})\| < \varepsilon_2 \|\nabla^P F(\mathbf{x}^{(0)}, \mathbf{h}^{(0)})\|$ ;
- maximum number of iteration:  $k + 1 \geq K$ .

---

**Algorithm 1** Modified PALM (mPALM) algorithm
 

---

Choose  $\mathbf{x}^{(0)} \in X$ ,  $\mathbf{h}^{(0)} \in H$ ,  $0 < \delta < 1$ ,  $\mu > 1$  and set  $\gamma_x = \gamma_h = 1$

```

for  $k = 0, 1, 2, \dots$  do
  1. Set  $c_k = \gamma_x L_x(\mathbf{h}^{(k)})$  and  $par = 1$ 
  2.  $\mathbf{x}^{(k+1)} = \mathcal{P}_X \left( \mathbf{x}^{(k)} - \frac{1}{c_k} \nabla_{\mathbf{x}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) \right)$ 
  3. Backtracking loop on  $c_k$ 
      if  $(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)})$  satisfies (4) and  $par = 1$  then
        Check for optimality condition on  $\mathbf{v}^{(k+1)} = \nabla_{\mathbf{x}}^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)})$ 
        if  $\|\mathbf{v}^{(k+1)}\| > \eta_x^{(k)}$  then
          set  $\gamma_x = \delta \gamma_x$  and go to step 1
        end if
      else
        set  $par = 2$ 
      end if
    Adjustment of  $c_k$ 
    if  $(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)})$  does not satisfy (4) and  $par = 2$  then
      set  $c_k = \mu \gamma_x L_x(\mathbf{h}^{(k)})$  and go to step 2
    end if
  4. Set  $\gamma_x = 1$  and define  $\eta_h^{(k+1)}$  according to (9).
  5. Set  $d_k = \gamma_h L_h(\mathbf{x}^{(k+1)})$  and  $par = 1$ 
  6.  $\mathbf{h}^{(k+1)} = \mathcal{P}_H \left( \mathbf{h}^{(k)} - \frac{1}{d_k} \nabla_{\mathbf{h}} F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) \right)$ 
  7. Backtracking loop on  $d_k$ 
      if  $(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})$  satisfies (5) and  $par = 1$  then
        Check for optimality condition on  $\mathbf{w}^{(k+1)} = \nabla_{\mathbf{h}}^P F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})$ 
        if  $\|\mathbf{w}^{(k+1)}\| > \eta_h^{(k+1)}$  then
          set  $\gamma_h = \delta \gamma_h$  and go to step 5
        end if
      else
        set  $par = 2$ 
      end if
    Adjustment of  $d_k$ 
    if  $(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)})$  does not satisfy (5) and  $par = 2$  then
      set  $d_k = \mu \gamma_h L_h(\mathbf{x}^{(k+1)})$  and go to step 6
    end if
  8. Set  $\gamma_h = 1$  and define  $\eta_x^{(k+1)}$  according to (8).
end for
  
```

---

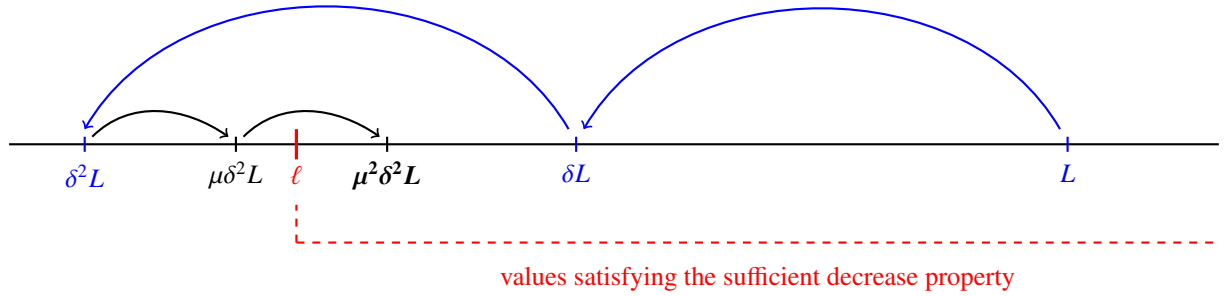


Figure 1: Example of procedure to determine the inverse of the steplength parameters. Here  $L$  denotes the initial guess while  $\ell$  represents the smaller parameter satisfying either (4) or (5).

### 3.2. Convergence properties

In this section we show the convergence property of the sequence generated by the modified PALM method to a critical point. The propositions and lemmas reported retrace the path followed in [6], suitable restricted to the considered blind deconvolution problem and extended to account for parameters  $c_k \in [c_{\min}, L_{\mathbf{x}}(\mathbf{h}^{(k)})]$ ,  $d_k \in [d_{\min}, L_{\mathbf{h}}(\mathbf{x}^{(k+1)})]$ , being  $c_{\min}, d_{\min}$  predetermined positive constants.

**Proposition 1** Assume that the function  $F$  in problem (2) belongs to  $\mathcal{F}$ . Then for all  $(\mathbf{x}, \mathbf{h}) \in \mathbb{R}^n \times \mathbb{R}^m$  we have

$$\begin{aligned}\partial\Psi(\mathbf{x}, \mathbf{h}) &= (\nabla_{\mathbf{x}}F(\mathbf{x}, \mathbf{h}) + \partial I_X(\mathbf{x}), \nabla_{\mathbf{h}}F(\mathbf{x}, \mathbf{h}) + \partial I_H(\mathbf{h})) = \\ &= (\partial_{\mathbf{x}}\Psi(\mathbf{x}, \mathbf{h}), \partial_{\mathbf{h}}\Psi(\mathbf{x}, \mathbf{h})).\end{aligned}$$

**Lemma 1 (Descent lemma)** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a continuously differentiable function with  $L_f$ -Lipschitz continuous gradient. Then

$$f(\mathbf{u}) \leq f(\mathbf{v}) + \langle \mathbf{u} - \mathbf{v}, \nabla f(\mathbf{v}) \rangle + \frac{L_f}{2} \|\mathbf{u} - \mathbf{v}\|^2 \quad \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d.$$

**Lemma 2 (Sufficient decrease property)** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a continuously differentiable function with  $L_f$ -Lipschitz continuous gradient,  $C \subseteq \mathbb{R}^d$  a closed, convex and nonempty set and  $\mathbf{u} \in C$ . Let  $t$  be a positive constant such that

$$f(\mathbf{u}^+) \leq f(\mathbf{u}) + \langle \mathbf{u}^+ - \mathbf{u}, \nabla f(\mathbf{u}) \rangle + \frac{t}{2} \|\mathbf{u}^+ - \mathbf{u}\|^2, \quad (10)$$

where  $\mathbf{u}^+$  is the projection of the vector  $\mathbf{u} - \frac{1}{t} \nabla f(\mathbf{u})$  onto  $C$ :

$$\mathbf{u}^+ = \mathcal{P}_C \left( \mathbf{u} - \frac{1}{t} \nabla f(\mathbf{u}) \right).$$

Then there exists  $t_1 > t$ ,  $A > 0$  such that

$$\langle \mathbf{u}^+ - \mathbf{u}, \nabla f(\mathbf{u}) \rangle \leq -\frac{t_1}{2} \|\mathbf{u}^+ - \mathbf{u}\|^2 < -\frac{t}{2} \|\mathbf{u}^+ - \mathbf{u}\|^2 \quad (11)$$

and a sufficient decrease in the value of  $f(\mathbf{u}^+)$  is ensured:

$$f(\mathbf{u}^+) < f(\mathbf{u}) - A^2. \quad (12)$$

*Proof.* First, we notice that it's always possible to find a parameter  $t$  as in (10) since the requested inequality for  $t$  is satisfied, at least, by selecting  $t = L_f$ .

From the definition of the projection onto a closed, convex and nonempty set  $C$ , the vector  $\mathbf{u}^+$  is characterized as

$$\mathbf{u}^+ = \arg \min_{\mathbf{v} \in C} \left\| \mathbf{u} - \frac{1}{t} \nabla f(\mathbf{u}) - \mathbf{v} \right\|^2$$

and in particular, by taking  $\mathbf{v} = \mathbf{u}$ , we obtain

$$\left\| \mathbf{u} - \frac{1}{t} \nabla f(\mathbf{u}) - \mathbf{u}^+ \right\|^2 < \left\| \mathbf{u} - \frac{1}{t} \nabla f(\mathbf{u}) - \mathbf{u} \right\|^2.$$

By writing the norm as scalar products one obtains the relation

$$\|\mathbf{u}^+ - \mathbf{u}\|^2 + \frac{2}{t} \langle \mathbf{u}^+ - \mathbf{u}, \nabla f(\mathbf{u}) \rangle < 0$$

which implies (11) by choosing e.g.  $t_1 = 2 \langle \mathbf{u}^+ - \mathbf{u}, \nabla f(\mathbf{u}) \rangle / \|\mathbf{u}^+ - \mathbf{u}\|^2$ . This condition combined with (10) leads to (12).  $\square$

For sake of simplicity, in the following we will use the notation  $\mathbf{z}^{(k)} = (\mathbf{x}^{(k)}, \mathbf{h}^{(k)})$  ( $k \in \mathbb{N}$ ).



**Lemma 3 (Convergence properties)** Suppose that  $F \in \mathcal{F}$  and let  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  be a sequence generated by the modified PALM method. The following assertions hold true:

- the sequence  $\{F(\mathbf{z}^{(k)})\}_{k \in \mathbb{N}}$  is nonincreasing and

$$\frac{\rho}{2} \|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|^2 \leq F(\mathbf{z}^{(k)}) - F(\mathbf{z}^{(k+1)}) \quad \forall k \geq 0, \quad (13)$$

where  $\rho = \min\{(c_k^1 - c_k), (d_k^1 - d_k)\}$  with  $c_k^1 > c_k$  and  $d_k^1 > d_k$  opportunely chosen;

- we have

$$\sum_{k=1}^{+\infty} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 + \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|^2 = \sum_{k=1}^{+\infty} \|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|^2 < +\infty$$

and, hence,  $\lim_{k \rightarrow +\infty} \|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\| = 0$ .

*Proof.* From condition (ii) of Definition 1, the function  $\mathbf{x} \rightarrow F(\mathbf{x}, \mathbf{h})$  (for  $\mathbf{h}$  fixed) is differentiable and has a  $L_x(\mathbf{h})$ -Lipschitz continuous gradient. Suppose that  $\mathbf{x}^{(k+1)}$  is computed as in Step 2 of Algorithm 1, with the parameter  $c_k$  satisfying condition (4). From Lemma 2 with  $f(\cdot) := F(\cdot, \mathbf{h}^{(k)})$  there exists  $c_k^1 > c_k$  such that

$$\begin{aligned} F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) &\leq F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) + \langle \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \nabla_{\mathbf{x}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) \rangle + \frac{c_k}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \\ &\leq F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) - \frac{c_k^1}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 + \frac{c_k}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \\ &= F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) + \frac{c_k - c_k^1}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2. \end{aligned}$$

In a similar way, one can prove that there exists  $d_k^1 > d_k$  such that

$$F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k+1)}) \leq F(\mathbf{x}^{(k+1)}, \mathbf{h}^{(k)}) + \frac{d_k - d_k^1}{2} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|^2.$$

Adding the above two inequalities, we obtain for all  $k \geq 0$  that

$$F(\mathbf{z}^{(k)}) - F(\mathbf{z}^{(k+1)}) \geq \frac{c_k^1 - c_k}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 + \frac{d_k^1 - d_k}{2} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|^2 \geq 0.$$

It follows that  $\{F(\mathbf{z}^{(k)})\}_{k \in \mathbb{N}}$  is a nonincreasing sequence bounded from below, and therefore it converges to some real number  $\bar{F}$ . Moreover, by choosing  $\rho = \min\{(c_k^1 - c_k), (d_k^1 - d_k)\}$ , we get

$$\begin{aligned} F(\mathbf{z}^{(k)}) - F(\mathbf{z}^{(k+1)}) &\geq \frac{\rho}{2} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 + \frac{\rho}{2} \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|^2 \\ &= \frac{\rho}{2} \|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|^2, \end{aligned}$$

and the first assertion is proved.

Let  $N$  be a positive integer. Summing (13) from  $k = 0$  to  $N - 1$  we also get

$$\begin{aligned} \sum_{k=0}^{N-1} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 + \|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|^2 &= \sum_{k=0}^{N-1} \|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|^2 \\ &\leq \frac{2}{\rho} (F(\mathbf{z}^{(0)}) - F(\mathbf{z}^{(N)})) \\ &\leq \frac{2}{\rho} (F(\mathbf{z}^{(0)}) - \bar{F}). \end{aligned}$$

By taking the limit as  $N \rightarrow +\infty$ , the second part of the lemma is established.  $\square$

**Lemma 4** Suppose that  $F \in \mathcal{F}$  and let  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  be a bounded sequence generated by the modified PALM method. For each positive integer  $k$ , define

$$\begin{aligned} A_{\mathbf{x}}^{(k)} &= c_{k-1}(\mathbf{x}^{(k-1)} - \mathbf{x}^{(k)}) + \nabla_{\mathbf{x}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) - \nabla_{\mathbf{x}} F(\mathbf{x}^{(k-1)}, \mathbf{h}^{(k-1)}), \\ A_{\mathbf{h}}^{(k)} &= d_{k-1}(\mathbf{h}^{(k-1)} - \mathbf{h}^{(k)}) + \nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) - \nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k-1)}). \end{aligned}$$

Then  $(A_{\mathbf{x}}^{(k)}, A_{\mathbf{h}}^{(k)}) \in \partial \Psi(\mathbf{x}^{(k)}, \mathbf{h}^{(k)})$  and there exists  $\zeta > 0$  such that

$$\|(A_{\mathbf{x}}^{(k)}, A_{\mathbf{h}}^{(k)})\| \leq \|A_{\mathbf{x}}^{(k)}\| + \|A_{\mathbf{h}}^{(k)}\| \leq \zeta \|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\|$$

for all  $k \geq 1$ .

*Proof.* Let  $k$  be a positive integer. From Step 2 of Algorithm 1 and recalling that the projection onto the set  $X$  can be seen as the proximal operator of the indicator function  $\mathcal{I}_X$ , we have

$$\mathbf{x}^{(k)} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \langle \mathbf{x} - \mathbf{x}^{(k-1)}, \nabla_{\mathbf{x}} F(\mathbf{x}^{(k-1)}, \mathbf{h}^{(k-1)}) \rangle + \frac{c_{k-1}}{2} \|\mathbf{x} - \mathbf{x}^{(k-1)}\|^2 + \mathcal{I}_X(\mathbf{x}) \right\}.$$

Writing down the optimality condition yields

$$\nabla_{\mathbf{x}} F(\mathbf{x}^{(k-1)}, \mathbf{h}^{(k-1)}) + c_{k-1}(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}) + \mathbf{u}^{(k)} = 0,$$

where  $\mathbf{u}^{(k)} \in \partial \mathcal{I}_X(\mathbf{x}^{(k)})$ . Hence

$$\nabla_{\mathbf{x}} F(\mathbf{x}^{(k-1)}, \mathbf{h}^{(k-1)}) + \mathbf{u}^{(k)} = c_{k-1}(\mathbf{x}^{(k-1)} - \mathbf{x}^{(k)}).$$

In a similar way, Step 6 of Algorithm 1 leads to

$$\mathbf{h}^{(k)} = \arg \min_{\mathbf{h} \in \mathbb{R}^m} \left\{ \langle \mathbf{h} - \mathbf{h}^{(k-1)}, \nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k-1)}) \rangle + \frac{d_{k-1}}{2} \|\mathbf{h} - \mathbf{h}^{(k-1)}\|^2 + \mathcal{I}_H(\mathbf{h}) \right\}$$

which means that

$$\nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k-1)}) + \mathbf{v}^{(k)} = d_{k-1}(\mathbf{h}^{(k-1)} - \mathbf{h}^{(k)}),$$

with  $\mathbf{v}^{(k)} \in \partial \mathcal{I}_H(\mathbf{h}^{(k)})$ . Since from Proposition 1 we have

$$\begin{aligned} \nabla_{\mathbf{x}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) + \mathbf{u}^{(k)} &\in \partial_{\mathbf{x}} \Psi(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}), \\ \nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) + \mathbf{v}^{(k)} &\in \partial_{\mathbf{h}} \Psi(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}), \end{aligned}$$

it follows that  $(A_{\mathbf{x}}^{(k)}, A_{\mathbf{h}}^{(k)}) \in \partial \Psi(\mathbf{x}^{(k)}, \mathbf{h}^{(k)})$ .

Let us prove now the second part of the lemma. Since  $\nabla F$  is locally Lipschitz continuous (see condition (iii) of Definition 1) and we assumed that  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  is bounded, there exists  $M > 0$  such that

$$\begin{aligned} \|A_{\mathbf{x}}^{(k)}\| &\leq c_{k-1} \|\mathbf{x}^{(k-1)} - \mathbf{x}^{(k)}\| + \|\nabla_{\mathbf{x}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) - \nabla_{\mathbf{x}} F(\mathbf{x}^{(k-1)}, \mathbf{h}^{(k-1)})\| \\ &\leq c_{k-1} \|\mathbf{x}^{(k-1)} - \mathbf{x}^{(k)}\| + M (\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| + \|\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}\|) \\ &= (M + c_{k-1}) \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| + M \|\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}\|. \end{aligned}$$

By definition, the constant  $c_{k-1}$  is less or equal to  $L_{\mathbf{x}}(\mathbf{h}^{(k-1)})$  and, from condition (ii) of Definition 1, the Lipschitz constants  $\{L_{\mathbf{x}}(\mathbf{h}^{(k-1)})\}_{k \in \mathbb{N}}$  are bounded from above by a fixed value  $\alpha_{\mathbf{x}}$ . As a result of this, we have

$$\begin{aligned} \|A_{\mathbf{x}}^{(k)}\| &\leq (M + \alpha_{\mathbf{x}}) \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| + M \|\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}\| \\ &\leq (2M + \alpha_{\mathbf{x}}) \|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\|. \end{aligned}$$

On the other hand, from the Lipschitz continuity of  $\nabla_{\mathbf{h}} F(\mathbf{x}, \cdot)$  and the values range for  $d_k$ , we have

$$\begin{aligned} \|A_{\mathbf{h}}^{(k)}\| &\leq d_{k-1} \|\mathbf{h}^{(k-1)} - \mathbf{h}^{(k)}\| + \|\nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) - \nabla_{\mathbf{h}} F(\mathbf{x}^{(k)}, \mathbf{h}^{(k-1)})\| \\ &\leq d_{k-1} \|\mathbf{h}^{(k-1)} - \mathbf{h}^{(k)}\| + L_{\mathbf{h}}^{(k)} \|\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}\| \\ &\leq 2L_{\mathbf{h}}^{(k)} \|\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}\|. \end{aligned}$$

Since  $\{L_h(\mathbf{x}^{(k)})\}_{k \in \mathbb{N}}$  is bounded from above by a constant  $\alpha_h$  (condition (ii) of Definition 1), the following inequality holds

$$\|A_h^{(k)}\| \leq 2\alpha_h \|\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}\| \leq 2\alpha_h \|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\|.$$

If we sum up the previous inequalities and define  $\tau = \max(\alpha_x, \alpha_h)$ , we obtain

$$\|(A_x^{(k)}, A_h^{(k)})\| \leq \|A_x^{(k)}\| + \|A_h^{(k)}\| \leq (2M + 3\tau) \|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\|$$

which leads to the thesis by choosing  $\zeta = (2M + 3\tau)$ .  $\square$

Let  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  be a sequence generated by the modified PALM from a starting point  $\mathbf{z}^{(0)} \in X \times H$ , and let  $\Omega(\mathbf{z}^{(0)})$  be the set of all its limit points, i.e.,

$$\Omega(\mathbf{z}^{(0)}) = \left\{ \bar{\mathbf{z}} \in \mathbb{R}^n \times \mathbb{R}^m : \exists K \subseteq \mathbb{N} \text{ such that } \lim_{k \in K, k \rightarrow \infty} \mathbf{z}^{(k)} = \bar{\mathbf{z}} \right\}.$$

Moreover, we denote with  $\text{crit}(f)$  the set of critical points of a function  $f$ . The following lemma and the convergence property in Theorem 1 can be demonstrated by following the same proofs in [6] and by exploiting the results previously shown in this section.

**Lemma 5** Suppose that  $F \in \mathcal{F}$  and let  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  be a bounded sequence generated by the modified PALM method. Then:

1.  $\emptyset \neq \Omega(\mathbf{z}^{(0)}) \subset \text{crit}(\Psi)$ ;
2.  $\lim_{k \rightarrow \infty} \text{dist}(\mathbf{z}^{(k)}, \Omega(\mathbf{z}^{(0)})) = 0$ ;
3.  $\Omega(\mathbf{z}^{(0)})$  is a nonempty, compact and connected set;
4. the objective function  $\Psi$  is finite and constant on  $\Omega(\mathbf{z}^{(0)})$ .

**Theorem 1** Suppose that  $\Psi$  is a KL function such that  $F \in \mathcal{F}$ , and let  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  be a bounded sequence generated by the modified PALM algorithm. The following assertions hold true.

1. The sequence  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  has finite length, that is,

$$\sum_{k=0}^{\infty} \|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\| < \infty.$$

2. The sequence  $\{\mathbf{z}^{(k)}\}_{k \in \mathbb{N}}$  converges to a critical point  $\mathbf{z}^* = (\mathbf{x}^*, \mathbf{h}^*)$  of  $\Psi$ .

## 4. Numerical experiments

In this section we test the effectiveness of the proposed modified PALM method in some synthetic blind deconvolution problems. Besides the PALM algorithm, we compare the results with two other methodologies, namely the gradient projection alternating minimization (GPAM) method and the alternating proximal gradient (APG) algorithm, which have been recently proposed in a general nonconvex optimization framework and applied to non-negative matrix factorization problems.

### 4.1. Gradient projection alternating minimization

In [17] the author proposed an inexact alternating scheme to minimize a smooth function subject to separable constraints. In particular, at each outer iteration the approximate solution of each subproblem is obtained by performing few steps of a gradient projection (GP) algorithm [18, 19, 20] exploiting a suitable alternation of the Barzilai–Borwein steplengths [21, 22]. Applications of this method to the blind deconvolution of astronomical images can be found in [23, 24]. The GPAM scheme rewritten for the regularized blind deconvolution problem (1) is detailed in Algorithm 2. For the choice of the inner iterations of the algorithm, we will use again conditions (7)–(9) on the decrease of the projected gradient norm. Moreover, in order to have a uniform stopping criterion we will arrest the GPAM method with the same strategy exploited in the mPALM case and described in section 3.1.

---

**Algorithm 2** Gradient projection alternating minimization (GPAM) method

---

Choose the starting point  $(\mathbf{x}^{(0)}, \mathbf{h}^{(0)}) \in X \times H$  and the upper bounds  $N_x, N_h \geq 1$  for the inner iterations numbers.

**for**  $k = 0, 1, \dots$  **do**

1. Compute  $\mathbf{x}^{(k+1)}$  with  $N_x^{(k)} \leq N_x$  GP iterations applied to subproblem

$$\min_{\mathbf{x} \in X} F(\mathbf{x}, \mathbf{h}^{(k)})$$

starting from the point  $\mathbf{x}^{(k)}$ .

2. Compute  $\mathbf{h}^{(k)}$  with  $N_h^{(k)} \leq N_h$  GP iterations applied to subproblem

$$\min_{\mathbf{h} \in H} F(\mathbf{x}^{(k+1)}, \mathbf{h})$$

starting from the point  $\mathbf{h}^{(k)}$ .

**end for**

---

#### 4.2. Alternating proximal gradient

As a further comparison, we tested also the APG method proposed in [16], whose scheme retraces the PALM method provided that the two projections of Steps 2 and 6 of Algorithm 1 are substituted with

$$\begin{aligned}\widehat{\mathbf{x}}^{(k+1)} &= \mathcal{P}_X \left( \widehat{\mathbf{x}}^{(k)} - \frac{1}{c_k} \nabla_{\mathbf{x}} F(\widehat{\mathbf{x}}^{(k)}, \mathbf{h}^{(k)}) \right), \\ \mathbf{h}^{(k+1)} &= \mathcal{P}_H \left( \widehat{\mathbf{h}}^{(k)} - \frac{1}{d_k} \nabla_{\mathbf{h}} F(\mathbf{x}^{(k+1)}, \widehat{\mathbf{h}}^{(k)}) \right),\end{aligned}$$

being

$$\begin{aligned}\widehat{\mathbf{x}}^{(k)} &= \mathbf{x}^{(k)} + \omega_{\mathbf{x}}^{(k)} (\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}), & \omega_{\mathbf{x}}^{(k)} &\geq 0, \\ \widehat{\mathbf{h}}^{(k)} &= \mathbf{h}^{(k)} + \omega_{\mathbf{h}}^{(k)} (\mathbf{h}^{(k)} - \mathbf{h}^{(k-1)}), & \omega_{\mathbf{h}}^{(k)} &\geq 0\end{aligned}$$

two suitable extrapolated points. Convergence of the sequence generated by the APG algorithm to a critical point of  $\Psi$  is again ensured by the KL property and the fact that  $F$  is a block-multiconvex function. As in the PALM case, we tried to exploit our backtracking strategy for the choice of  $c_k, d_k$  also to this method but without obtaining significant improvements. As a result of this, in our numerical tests we will use the Lipschitz constant of the partial gradients as in the PALM method. As concerns the extrapolation weights, we defined

$$\begin{aligned}\omega_{\mathbf{x}}^{(k)} &= \min \left( \frac{t_k - 1}{t_{k+1}}, 0.99 \sqrt{\frac{L_{\mathbf{x}}(\mathbf{h}^{(k-1)})}{L_{\mathbf{x}}(\mathbf{h}^{(k)})}} \right), \\ \omega_{\mathbf{h}}^{(k)} &= \min \left( \frac{t_k - 1}{t_{k+1}}, 0.99 \sqrt{\frac{L_{\mathbf{h}}(\mathbf{x}^{(k)})}{L_{\mathbf{h}}(\mathbf{x}^{(k+1)})}} \right),\end{aligned}$$

where [25, 26]

$$t_0 = 1, \quad t_k = \frac{1}{2} \left( 1 + \sqrt{1 + 4t_{k-1}^2} \right), \quad k = 1, 2, \dots$$

The stopping criterion is again the one used for the other approaches.

#### 4.3. Results

All the described methods have been tested in the following three simulated problems:

1. the modified Shepp–Logan *phantom* test image from the Matlab Image Processing toolbox, artificially blurred with a Gaussian PSF with standard deviation equal to 4. The regularization terms we chose in this case are  $R^{\text{HS}}$  for the image and  $R^{\text{T0}}$  for the PSF;
2. the *satellite* image frequently used in several papers on image deconvolution (see e.g. [27, 28, 29]), artificially blurred with an out-of-focus PSF with radius equal to 4. The regularization term we chose in this case is  $R^{\text{HS}}$  for both the image and the PSF;
3. the Hubble Space Telescope (HST) image of the *crab nebula* NGC 19521, artificially blurred with an Airy function [30] mimicking the ideal acquisition of one mirror of the Large Binocular Telescope (LBT - <http://www.lbto.org>). The regularization term we chose in this case is  $R^{\text{T0}}$  for both the image and the PSF.

The three blurred images have been corrupted with 5% Gaussian noise and the resulting data are shown in Figure 2 together with the true images  $\mathbf{x}^*$  and PSFs  $\mathbf{h}^*$ . All the images and the PSFs are sized  $256 \times 256$ . The initialization  $\mathbf{x}^{(0)}$  for the image has been set equal to the datum  $\mathbf{g}$  for all the datasets, while for the PSF we used as  $\mathbf{h}^{(0)}$  a Gaussian function with standard deviation equal to 5.5 for the phantom dataset, an out-of-focus PSF with radius equal to 5.5 for the satellite dataset and a Lorentzian function with half-width at half-maximum equal to 6 for the crab nebula dataset. The parameters for the tuning of  $c_k, d_k$  and the thresholds defining the stopping criteria have been chosen as follows:  $\delta = 0.1, \mu = 1.5, c_{\min} = d_{\min} = 10^{-10}, \varepsilon_1 = 10^{-14}, \varepsilon_2 = 10^{-6}, K = 1000$ . Finally, for each problem we manually selected the regularization parameters  $\lambda_1, \lambda_2$  which provided in average the best reconstructions. The numerical experiments have been performed by means of routines implemented by ourselves in Matlab R2014a and run on a PC equipped with a 1.80 GHz Intel(R) Core(TM) i7-4500U in a Windows 8 environment.

In Table 1 we reported the relative Euclidean reconstruction errors on image and PSF

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\|/\|\mathbf{x}^*\| \quad ; \quad \|\mathbf{h}^{(k)} - \mathbf{h}^*\|/\|\mathbf{h}^*\| \quad (14)$$

obtained by the GPAM, AGP, PALM and modified PALM methods, together with the number of iterations  $k$  performed to satisfy the stopping criterion and the computational time in seconds. For sake of completeness, we also added the errors (14) between the initial ( $k = 0$ ) and the true images and PSFs, and the best error achieved in a standard deconvolution framework by using the true and the initial PSF (the first value should represent the best error achievable by a blind deconvolution approach, while the second value should be a reference value that a blind deconvolution method has to improve to justify the heaviest computations of the blind scheme). The optimization algorithm we used to compute the nonblind reconstructions is the scaled gradient projection (SGP) method, proposed by Bonettini et al. [18] and applied in several image denoising and deblurring problems [31, 32, 33]. Moreover, in Figure 3 we plot the errors behavior as a function of the iteration numbers, while in Figure 4 the deconvolved images for the satellite dataset are shown.

Table 1: Results obtained with the considered algorithms in the three simulated datasets. For each problem, we show the errors on image ( $\text{err}_x$ ) and PSF ( $\text{err}_h$ ), the number of iterations required to satisfy the stopping criterion ( $\text{iter}$ ) and the execution time in seconds ( $\text{sec}$ ). Moreover, we also report the errors between true and initial images ( $\mathbf{e}_{x^{(0)}}$ ) and PSFs ( $\mathbf{e}_{h^{(0)}}$ ), together with the reconstruction errors on the image obtained in a standard deconvolution framework by using the true ( $S_I$ ) or the initial ( $S_0$ ) PSFs.

Phantom					Satellite				Crab nebula			
$\mathbf{e}_{x^{(0)}}: 0.666, \mathbf{e}_{h^{(0)}}: 0.381$					$\mathbf{e}_{x^{(0)}}: 0.566, \mathbf{e}_{h^{(0)}}: 0.704$				$\mathbf{e}_{x^{(0)}}: 0.393, \mathbf{e}_{h^{(0)}}: 0.977$			
$S_I: 0.384, S_0: 0.455$					$S_I: 0.192, S_0: 0.275$				$S_I: 0.176, S_0: 0.225$			
method	iter	sec	$\text{err}_x$	$\text{err}_h$	iter	sec	$\text{err}_x$	$\text{err}_h$	iter	sec	$\text{err}_x$	$\text{err}_h$
GPAM	21	07.2	0.398	0.106	35	13.6	0.235	0.418	95	32.2	0.186	0.419
AGP	252	12.9	0.399	0.097	457	25.9	0.217	0.387	1000	47.8	0.201	0.588
PALM	1000	45.3	0.453	0.278	1000	51.0	0.410	0.636	1000	38.7	0.264	0.958
mPALM	125	21.2	0.395	0.090	117	21.4	0.218	0.395	183	37.2	0.182	0.372

From the numerical results and the error plots we can do the following considerations:

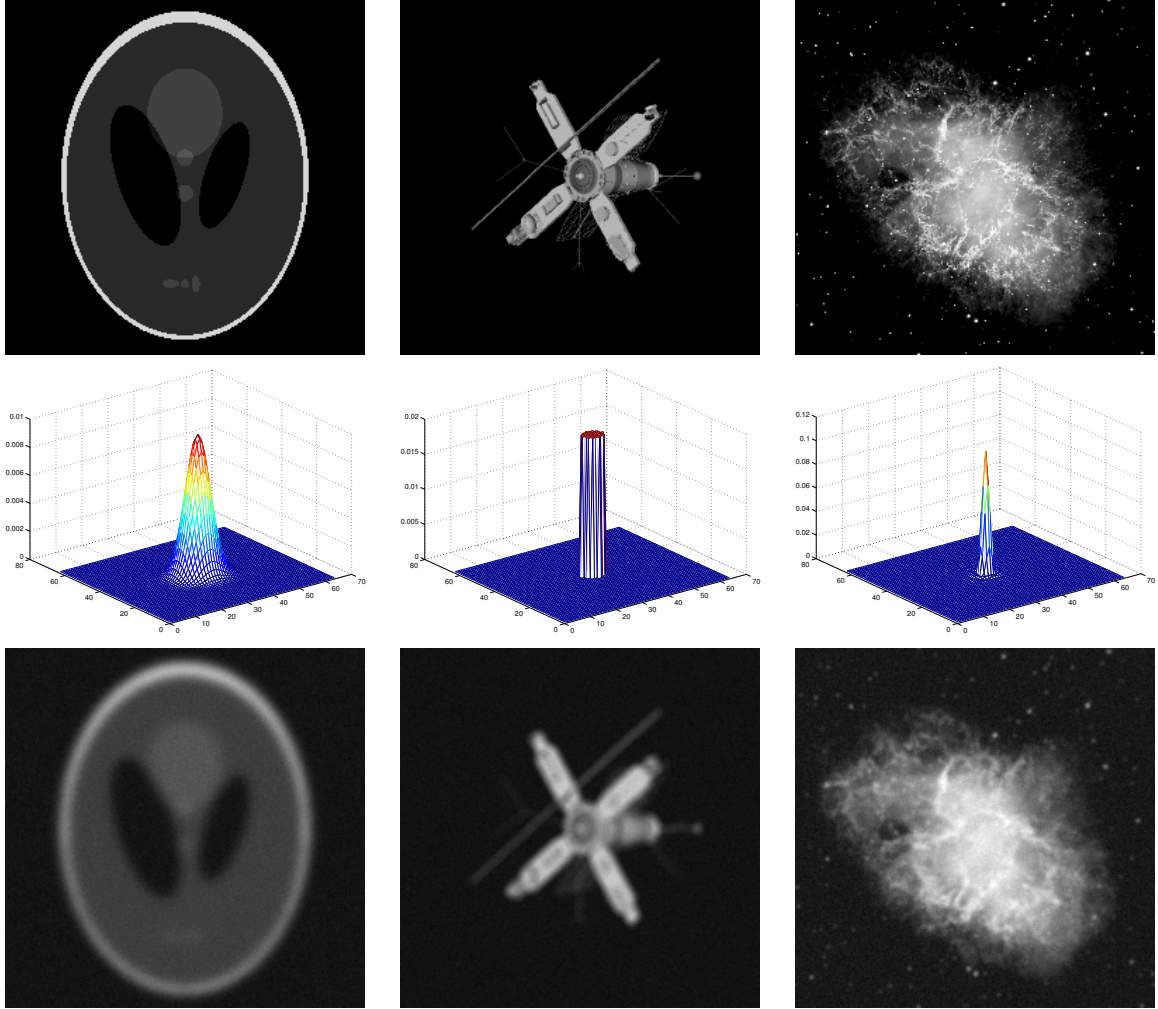


Figure 2: True images (top row) and PSFs (central row - zoom on the  $64 \times 64$  central panel) used for the numerical tests. The blurred and noisy images are also shown (bottom row).

- the backtracking procedure in the choice of the steplengths strongly improves the convergence of the PALM method, with the result that good reconstructions can be achieved in a computational time which is comparable to those of the GPAM and AGP approaches. Here the GPAM method provides higher errors in a lower time, but this is mainly due to the stopping criteria we adopted, since for this method the graphs of the errors are still decreasing (see Figure 3). In further tests with different thresholds  $\varepsilon_1, \varepsilon_2$  we verified that both the errors and the computational times of GPAM become closer to those of mPALM;
- the reconstruction errors achieved by the blind schemes are very close to the optimal ones obtained in a standard deconvolution framework with the true PSF, and significantly lower than the analogous ones provided by choosing an approximation of the PSF (in our case,  $\mathbf{h}^{(0)}$ ), thus confirming the potential benefit of the blind approach;
- the lower number of iterations observed in the mPALM case with respect to the PALM and AGP methods, for which the convergence to a critical point has been proved, gives hope that a further improvement of the backtracking procedure which requires fewer steps (i.e., fewer fast Fourier transform computations) can lead to a significant reduction of the whole computational effort. We tried with a kind of “hot start” of the

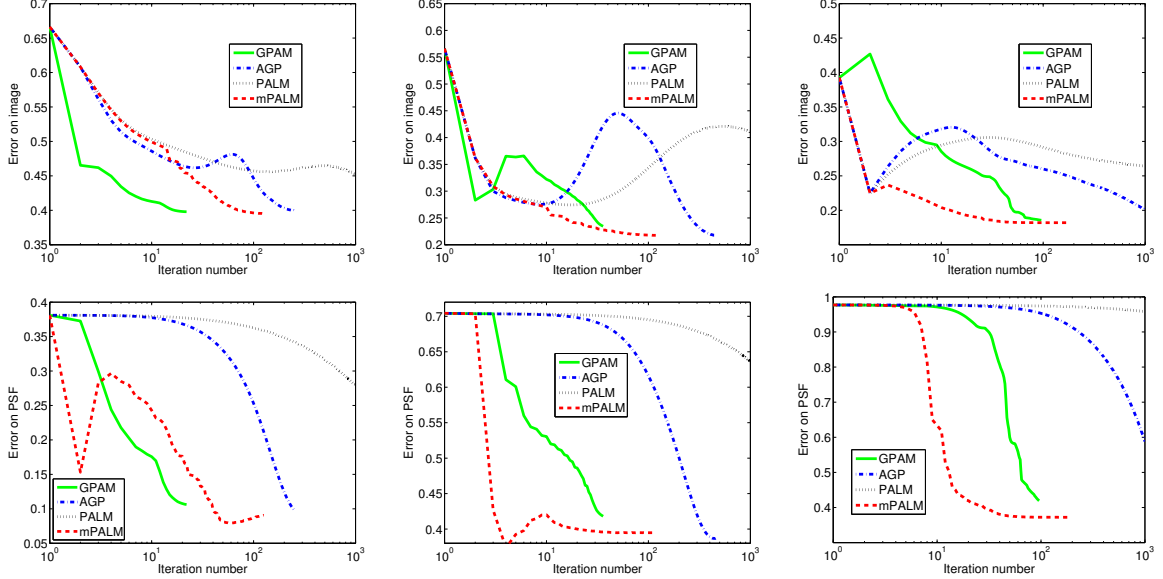


Figure 3: Plots of the reconstruction errors on the image (top row) and the PSF (bottom row) as functions of the iteration number for the phantom (left column), satellite (central column) and crab nebula (right column) datasets.

parameters  $\gamma_x, \gamma_h$  by using as initializations the values obtained at the previous iteration, as done e.g. in [34, 35], but we did not find any significant improvement. Further tests with means over a limited number of previous values and/or different (and possibly adaptive) factors  $\delta, \mu$  will be investigated in a future work.

## 5. Conclusions

In this paper we address the blind deconvolution problem with images perturbed by Gaussian noise and we consider a recent approach for a more general class of nonconvex and nonsmooth minimization problems whose sequence of iterates converges to a critical point. Our work has been to exploit the same arguments which led to the convergence proof of the method and extend it to the case of wider ranges of parameters, provided that the regularization terms used in the formulation of the blind deconvolution problems satisfy certain properties. The possibility to select steplength larger than the inverse of the Lipschitz constants of the partial gradients allows a faster convergence of the method to the critical point, and we proposed a backtracking procedure to automatically select these parameters based on the optimality conditions of the problem. Some numerical experiments with different images, PSFs and combinations of regularization terms showed that the new variation of the method strongly reduces the number of iterations required by the PALM algorithm, and provides comparable results with respect to other state-of-the-art methods, leaving also space for further improvements.

A future development of this work will be the application of the proposed approach on real image deconvolution problems in astronomy, as the reconstruction of X-ray emission from a solar flare [23, 36], in which a more detailed reconstructed image can strongly help in detecting some crucial features of the phenomenon under investigation (see e.g. [37, 38]).

## Acknowledgments

This work has been partially supported by MIUR (Italian Ministry for University and Research), under the two projects FIRB – Futuro in Ricerca 2012, Contract RBFR12M3AC, and PRIN 2012, Contract 2012MTE38N, and by the Italian Spinner2013 PhD Project “High-complexity inverse problems in biomedical applications and social systems”. The Italian GNCS – INdAM (Gruppo Nazionale per il Calcolo Scientifico – Istituto Nazionale di Alta Matematica) is also acknowledged.



Figure 4: Image reconstructed with the GPAM (top left panel), AGM (top right panel), PALM (bottom left panel) and mPALM (bottom right panel) in the satellite case.



## References

- [1] M. Bertero, P. Boccacci, Introduction to inverse problems in imaging, Institute of Physics Publishing, Bristol, 1998.
- [2] M. S. C. Almeida, M. A. T. Figueiredo, Parameter estimation for blind and non-blind deblurring using residual whiteness measures, *IEEE T. Image Process.* 22 (7) (2013) 2751–2763.
- [3] J. P. Oliveira, M. A. T. Figueiredo, J. M. Bioucas-Dias, Parametric blur estimation for blind restoration of natural images: linear motion and out-of-focus, *IEEE T. Image Process.* 23 (1) (2014) 466–477.
- [4] P. L. Combettes, V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.* 4 (4) (2005) 1168–1200.
- [5] P. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in: H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, H. Wolkowicz (Eds.), *Fixed-point algorithms for inverse problems in science and engineering*, Springer Optimization and Its Applications, Springer, New York, NY, 2011, pp. 185–212.
- [6] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.* 146 (1–2) (2014) 459–494.
- [7] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, in: *Les Équations aux Dérivées Partielles*, Éditions du Centre National de la Recherche Scientifique, Paris, 1963, pp. 87–89.
- [8] S. Łojasiewicz, Sur la géométrie semi- et sous-analytique, *Ann. Inst. Fourier* 43 (5) (1993) 1575–1595.
- [9] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* 48 (3) (1998) 769–783.
- [10] P. C. Hansen, J. G. Nagy, D. P. O’Leary, *Deblurring Images: Matrices, Spectra and Filtering*, SIAM, Philadelphia, 2006.
- [11] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *J. Math. Imaging Vis.* 40 (2011) 120–145.
- [12] T. L. Jensen, J. Jørgensen, P. Hansen, S. H. Jensen, Implementation of an optimal first-order method for strongly convex total variation regularization, *BIT* 52 (2) (2012) 329–356.
- [13] D. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, 1999.
- [14] M. Razaviyayn, M. Hong, Z.-Q. Luo, A unified convergence analysis of block successive minimization methods for nonsmooth optimization, *SIAM J. Optimiz.* 23 (2) (2013) 1126–1153.
- [15] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality, *Math. Oper. Res.* 35 (2) (2010) 438–457.
- [16] Y. Xu, W. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, *SIAM J. Imaging Sci.* 6 (3) (2013) 1758–1789.
- [17] S. Bonettini, Inexact block coordinate descent methods with application to the nonnegative matrix factorization, *IMA J. Numer. Anal.* 31 (4) (2011) 1431–1452.
- [18] S. Bonettini, R. Zanella, L. Zanni, A scaled gradient projection method for constrained image deblurring, *Inverse Probl.* 25 (1) (2009) 015002.
- [19] S. Bonettini, M. Prato, Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm, *Inverse Probl.* 26 (9) (2010) 095001.
- [20] M. Prato, R. Cavicchioli, L. Zanni, P. Boccacci, M. Bertero, Efficient deconvolution methods for astronomical imaging: algorithms and IDL-GPU codes, *Astron. Astrophys.* 539 (2012) A133.
- [21] J. Barzilai, J. M. Borwein, Two-point step size gradient methods, *IMA J. Numer. Anal.* 8 (1) (1988) 141–148.
- [22] G. Frassoldati, G. Zanghirati, L. Zanni, New adaptive stepsize selections in gradient methods, *J. Ind. Manage. Optim.* 4 (2) (2008) 299–312.
- [23] S. Bonettini, A. Cornelio, M. Prato, A new semiblind deconvolution approach for Fourier-based image restoration: an application in astronomy, *SIAM J. Imaging Sci.* 6 (3) (2013) 1736–1757.
- [24] M. Prato, A. La Camera, S. Bonettini, M. Bertero, A convergent blind deconvolution method for post-adaptive-optics astronomical imaging, *Inverse Probl.* 29 (6) (2013) 065017.
- [25] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* 2 (1) (2009) 183–202.
- [26] Y. Nesterov, A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ , *Doklady Akad. Nauk SSSR* 27 (1983) 372–376.
- [27] M. S. C. Almeida, L. B. Almeida, Blind and semi-blind deblurring of natural images, *IEEE T. Image Process.* 19 (1) (2010) 36–52.
- [28] P. Brianzi, F. Di Benedetto, C. Estatico, Preconditioned iterative regularization in banach spaces, *Comput. Optim. Appl.* 54 (2) (2013) 263–282.
- [29] A. Staglianò, P. Boccacci, M. Bertero, Analysis of an approximate model for poisson data reconstruction and a related discrepancy principle, *Inverse Probl.* 27 (12) (2011) 125003.
- [30] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 1972.
- [31] F. Benvenuto, R. Zanella, L. Zanni, M. Bertero, Nonnegative least-squares image deblurring: improved gradient projection approaches, *Inverse Probl.* 26 (2) (2010) 025004.
- [32] R. Zanella, P. Boccacci, L. Zanni, M. Bertero, Efficient gradient projection methods for edge-preserving removal of Poisson noise, *Inverse Probl.* 25 (4) (2009) 045010.
- [33] R. Zanella, G. Zanghirati, R. Cavicchioli, L. Zanni, P. Boccacci, M. Bertero, G. Vicidomini, Towards real-time image deconvolution: application to confocal and sted microscopy, *Sci. Rep.* 3 (2013) 2523.
- [34] C. J. Lin, Projected gradient methods for nonnegative matrix factorization, *Neural Comput.* 19 (10) (2007) 2756–2779.
- [35] C. J. Lin, J. J. Moré, Newton’s method for large-scale bound constrained problems, *SIAM J. Optimiz.* 9 (4) (1999) 1100–1127.
- [36] M. Prato, Regularization methods for the solution of inverse problems in solar X-ray and imaging spectroscopy, *Arch. Comput. Method E.* 16 (2) (2009) 109–160.

- [37] M. Prato, A. G. Emslie, E. P. Kontar, A. M. Massone, M. Piana, The location of centroids in photon and electron maps of solar flares, *Astrophys. J.* 706 (1) (2009) 917–922.
- [38] Y. Xu, A. G. Emslie, G. J. Hurford, RHESSI hard X-ray imaging spectroscopy of extended sources and the physical properties of electron acceleration regions in solar flares, *Astrophys. J.* 673 (1) (2008) 576–585.