

This is a pre print version of the following article:

Application of cyclic block generalized gradient projection methods to Poisson blind deconvolution / Rebegoldi, Simone; Bonettini, Silvia; Prato, Marco. - ELETTRONICO. - (2015), pp. 225-229. (Intervento presentato al convegno 23rd European Signal Processing Conference, EUSIPCO 2015 tenutosi a Nice nel Aug. 31-Sept. 4 2015) [10.1109/EUSIPCO.2015.7362378].

IEEE Institute of Electrical and Electronics Engineers Inc.

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.

24/04/2024 06:53

APPLICATION OF CYCLIC BLOCK GENERALIZED GRADIENT PROJECTION METHODS TO POISSON BLIND DECONVOLUTION

*S. Rebegoldi**, *S. Bonettini†*, *M. Prato**

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

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

ABSTRACT

The aim of this paper is to consider a modification of a block coordinate gradient projection method with Armijo linesearch along the descent direction in which the projection on the feasible set is performed according to a variable non Euclidean metric. The stationarity of the limit points of the resulting scheme has recently been proved under some general assumptions on the generalized gradient projections employed. Here we tested some examples of methods belonging to this class on a blind deconvolution problem from data affected by Poisson noise, and we illustrate the impact of the projection operator choice on the practical performances of the corresponding algorithm.

Index Terms— Constrained optimization, gradient projection methods, alternating algorithms, nonconvex optimization.

1. INTRODUCTION

Block descent methods are useful when one addresses a general optimization problem

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}), \quad (1)$$

in which the constraint set, as in several relevant applications, has a separable structure, i.e. $\Omega = \Omega_1 \times \dots \times \Omega_m$, with $\Omega_i \subseteq \mathbb{R}^{n_i}$, $\sum_{i=1}^m n_i = n$, so that any $\mathbf{x} \in \Omega$ can be block partitioned as $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$.

Such methods are based on the idea of performing successive minimizations over each block, as in the classical nonlinear Gauss-Seidel method [1]:

$$\mathbf{x}_i^{(k+1)} \in \arg \min_{\mathbf{x} \in \Omega_i} f(\mathbf{x}_1^{(k+1)}, \dots, \mathbf{x}_{i-1}^{(k+1)}, \mathbf{x}, \mathbf{x}_{i+1}^{(k)}, \dots, \mathbf{x}_m^{(k)}). \quad (2)$$

This work has been partially supported by MIUR 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 is also acknowledged.

However, the convergence of this approach is not ensured without quite restrictive convexity assumptions (see [1, 2]) and, in addition, computing an exact minimum of f , even if restricted to a single block, can be impractical.

On the other side, inspired by the idea of (2), effective methods able to handle general nonconvex problems and with global convergence properties can be designed [3, 4]. In particular, in [3] the author proposed a cyclic block gradient projection method, based on inexact solutions of subproblems (2) obtained by exploiting the sufficient decrease property of the Armijo linesearch [1]. In a recent paper [5], this approach has been further developed, allowing generalized projections with variable parameters choice. In this work we consider three instances of projections belonging to this class and we apply the resulting schemes to a large-scale problem in astronomy, namely the blind deconvolution of stellar images acquired by a ground-based telescope. Future work will include the convergence analysis of the proposed scheme for Kurdyka-Łojasiewicz functions [6–8].

2. CYCLIC BLOCK GENERALIZED GRADIENT PROJECTION METHOD

We begin this section by defining the set $\mathcal{D}(\Omega)$ of the functions $d : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$ which are convex, continuously differentiable and such that

$$d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y} \quad (3)$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$. Given an array of parameters $\boldsymbol{\sigma} \in S \subseteq \mathbb{R}^q$ and a function $d_{\boldsymbol{\sigma}} \in \mathcal{D}(\Omega)$, we denote by $\mathcal{H}(f, \Omega, S)$ the set of the metric functions defined as

$$h_{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}) = \nabla f(\mathbf{y})^T (\mathbf{x} - \mathbf{y}) + d_{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}) \quad (4)$$

and, for any $h_{\boldsymbol{\sigma}} \in \mathcal{H}(f, \Omega, S)$, we define the associated generalized gradient projection operator $p(\cdot; h_{\boldsymbol{\sigma}}) : \Omega \rightarrow \Omega$ as

$$p(\mathbf{x}; h_{\boldsymbol{\sigma}}) = \arg \min_{\mathbf{z} \in \Omega} h_{\boldsymbol{\sigma}}(\mathbf{z}, \mathbf{x}) \quad \forall \mathbf{x} \in \Omega. \quad (5)$$

Examples of metric functions which can be rewritten in the form (4) are:

- a) the standard Euclidean projection $p(x; h_\sigma) = P_\Omega(x - \sigma \nabla f(x))$, obtained by choosing

$$d_\sigma(x, y) = \frac{1}{2\sigma} \|x - y\|^2, \quad \sigma > 0; \quad (6)$$

- b) the scaled Euclidean projection, considered for example in [9, 10], corresponding to the choice

$$d_{(\alpha, D)}(x, y) = \frac{1}{2\alpha} (x - y)^T D^{-1} (x - y). \quad (7)$$

In this case the array of parameters σ is given by the pair (α, D) , where $\alpha \in \mathbb{R}_{>0}$ and $D \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix;

- c) the Bregman distance associated to a strictly convex function $b : \Omega \rightarrow \mathbb{R}$, which is defined as

$$d_\sigma(x, y) = \frac{1}{\sigma} (b(x) - b(y) - \nabla b(y)^T (x - y)), \quad \sigma > 0. \quad (8)$$

In order to formally introduce our method for problem (1), we choose the metric function $h_\sigma \in \mathcal{H}(f, \Omega, S)$, where $S = S_1 \times \dots \times S_m$, $S_i \subset \mathbb{R}^{q_i}$, such that the parameter σ can be partitioned as $\sigma = (\sigma_1, \dots, \sigma_m)$. Moreover, we define h_σ so that it is separable over the m blocks with respect to its first variable, i.e.

$$\begin{aligned} h_\sigma(x, y) &= \sum_{i=1}^m h_{\sigma_i}^i(x_i, y) \\ &= \sum_{i=1}^m \nabla_i f(y)^T (x_i - y_i) + d_{\sigma_i}^i(x_i, y_i), \end{aligned} \quad (9)$$

where the functions $d_{\sigma_i}^i$ belong to $\mathcal{D}(\Omega_i)$ ($i = 1, \dots, m$) and are chosen e.g. as in (6)–(8). It is easy to see that the metric function h_σ defined in (9) belongs to $\mathcal{H}(f, \Omega, S)$ and the associated generalized gradient projection can be also partitioned by blocks as

$$p(x; h_\sigma) = \begin{pmatrix} p_1(x; h_{\sigma_1}^1) \\ \vdots \\ p_m(x; h_{\sigma_m}^m) \end{pmatrix},$$

where

$$p_i(x; h_{\sigma_i}^i) = \arg \min_{z_i \in \Omega_i} h_{\sigma_i}^i(z_i, x).$$

The resulting cyclic block generalized gradient projection method is outlined in Algorithm 1.

The convergence result for this scheme has been proved in [5]. For sake of completeness, we report the statement of the main theorem.

Theorem 1 *Let $\{x^{(k)}\}_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1 and assume that \bar{x} is a limit point of $\{x^{(k)}\}_{k \in \mathbb{N}}$. Then \bar{x} is a stationary point for problem (1).*

Algorithm 1 Cyclic block generalized gradient projection method

Define a compact set S and a metric function $h_\sigma \in \mathcal{H}(f, \Omega, S)$ as in (9). Choose $x^{(0)} \in \Omega$, the upper bounds for the inner iterations numbers L_1, \dots, L_m and $\beta, \delta \in (0, 1)$. FOR $k = 0, 1, 2, \dots$

FOR $i = 1, \dots, m$

Set $x_i^{(k,0)} = x_i^{(k)}$

Choose the inner iterations number $L_i^{(k)} \leq L_i$

FOR $\ell = 0, \dots, L_i^{(k)} - 1$

Choose the parameter $\sigma_i^{(k,\ell)} \in S_i$

Set $\tilde{x}^{(k,\ell)} = (x_1^{(k+1)}, \dots, x_{i-1}^{(k+1)}, x_i^{(k,\ell)}, \dots, x_m^{(k)})$

Compute the descent direction

$$d_i^{(k,\ell)} = p_i \left(\tilde{x}^{(k,\ell)}; h_{\sigma_i^{(k,\ell)}}^i \right) - x_i^{(k,\ell)}$$

Set $\tilde{d}^{(k,\ell)} = (0, \dots, 0, d_i^{(k,\ell)}, \dots, 0)$

Compute the linesearch parameter $\lambda_i^{(k,\ell)} = \delta^{m_{k,l}}$, where $m_{k,l}$ is the smallest non-negative integer such that

$$\begin{aligned} f(\tilde{x}^{(k,\ell)} + \lambda_i^{(k,\ell)} \tilde{d}^{(k,\ell)}) &\leq f(\tilde{x}^{(k,\ell)}) \\ &\quad + \beta \lambda_i^{(k,\ell)} \nabla f(\tilde{x}^{(k,\ell)})^T \tilde{d}^{(k,\ell)} \end{aligned}$$

Set $x_i^{(k,\ell+1)} = x_i^{(k,\ell)} + \lambda_i^{(k,\ell)} d_i^{(k,\ell)}$

END

Set $x_i^{(k+1)} = x_i^{(k, L_i^{(k)})}$

END

END

3. POISSON BLIND DECONVOLUTION

In the Poisson image blind deconvolution problem, the basic assumption is that the available data $g \in \mathbb{R}^n$ is a realization of a Poisson random variable whose mean is $\omega^* \otimes x^* + be$, where $\omega^* \in \mathbb{R}^n$ is an unknown point spread function (PSF), \otimes denotes the convolution operator (periodic boundary conditions are assumed here), b is a positive parameter representing the background radiation, $e \in \mathbb{R}^n$ is the vector of all ones and $x^* \in \mathbb{R}^n$ is the image we would like to recover. In the following, we will assume that the PSF ω^* is normalized to one. Following a maximum likelihood approach, one can consider the optimization problem

$$\min_{x \in \Omega_x, \omega \in \Omega_\omega} KL(x, \omega), \quad (10)$$

where $KL(\mathbf{x}, \boldsymbol{\omega})$ is the Kullback–Leibler divergence

$$KL(\mathbf{x}, \boldsymbol{\omega}) = \sum_{i=1}^n g_i \log \left(\frac{g_i}{(\boldsymbol{\omega} \otimes \mathbf{x})_i + b} \right) + (\boldsymbol{\omega} \otimes \mathbf{x})_i + b - g_i, \quad (11)$$

and introduce regularization by solving it approximately through the early stopping of an iterative procedure.

As concerns the feasible sets, we consider non-negativity and flux conservation for the image \mathbf{x} , while for the PSF $\boldsymbol{\omega}$ we impose non-negativity, normalization to 1 and an upper bound s which can be estimated in the case of adaptive optics devices from the knowledge of the so-called Strehl ratio (SR), i.e. the ratio of peak diffraction intensity of an aberrated versus perfect waveform (see e.g. [11]). These constraints in a blind deconvolution framework have been used e.g. in [12–14] and allow good reconstructions even in presence of a large scale nonconvex problem as (10). The resulting sets are then given by

$$\Omega_{\mathbf{x}} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq 0, \sum_{i=1}^n x_i = \sum_{i=1}^n g_i - nb\},$$

$$\Omega_{\boldsymbol{\omega}} = \{\boldsymbol{\omega} \in \mathbb{R}^n \mid 0 \leq \boldsymbol{\omega} \leq s, \sum_{i=1}^n \omega_i = 1\}.$$

We consider a realistic simulation in the astronomical field by a) generating a 512×512 image \mathbf{x}^* of a cluster of 100 stars with different magnitudes (brightest value $\approx 3.2 \cdot 10^7$, dimmest value $\approx 4.2 \cdot 10^6$); b) convolving it with a PSF $\boldsymbol{\omega}^*$ (SR = 0.81) mimicking the response of a single mirror of the large binocular telescope (LBT); c) adding a realistic background radiation ($b \approx 2.6 \cdot 10^4$); and d) corrupting the resulting blurred image with Poisson noise. In Figure 1 we report both the PSF used in this experiment and the simulated measured image. The reconstruction algorithms are instances of Algorithm 1 where the projection is defined by means of the three choices detailed in section 2, whose main features are detailed below.

Scaled gradient projection (SGP). The scaled Euclidean projection (7) has been proposed in [10] within the so-called scaled gradient projection method and allowed a notable acceleration of the same method employing the standard Euclidean projection (6), as remarked in several recent papers [13, 15–21]. Here we adopt, at each iteration k , the following diagonal scaling matrix

$$[D_k]_{ii} = \max \left\{ \frac{1}{\mu}, \min \left\{ \mu, \mathbf{x}^{(k)} \right\} \right\}, \quad (12)$$

where μ is a prefixed threshold. The steplength parameter α_k is then computed by the adaptive alternation of the scaled Barzilai–Borwein (BB) rules as proposed in [10]. If the scaling matrix is set equal to the identity we recover the usual gradient projection (GP) method.

Gradient method with Bregman projection (GBP). A further instance of projections belonging to the family described in section 2 corresponds to the choice of the metric (4) with related distance-like function d_{σ} defined in (8), where $b(\mathbf{x}) = \sum_{i=1}^n (x_i + \gamma) \log(x_i + \gamma)$, with $\gamma > 0$, is the “regularized entropy” [22]. The resulting projection operator (5) is given by

$$[\mathbf{p}(\mathbf{x}, h_{\sigma})]_i = \max \left\{ (x_i + \gamma) e^{-\sigma \nabla_i f(\mathbf{x})} - \gamma, 0 \right\}.$$

The steplength parameter σ is adaptively computed at each iteration in the following way. First, we observe that, by the Taylor expansion of the exponential function, we have

$$(x_i + \gamma) e^{-\sigma \nabla_i f(\mathbf{x})} = (x_i + \gamma) - q_i(\sigma)(x_i + \gamma) \nabla_i f(\mathbf{x}),$$

where $q_i(\sigma) = \sum_{j=0}^{\infty} (-1)^j \frac{\sigma^{j+1}}{(j+1)!} \nabla_i f(\mathbf{x})^j$. The term $q_i(\sigma)$ can be explicitly expressed also as $q_i(\sigma) = (1 - e^{-\sigma \nabla_i f(\mathbf{x})}) / \nabla_i f(\mathbf{x})$ when $\nabla_i f(\mathbf{x}) \neq 0$, $q_i(\sigma) = \sigma$ when $\nabla_i f(\mathbf{x}) = 0$. Then, the GBP method can be considered also as an approximated scaled gradient method employing the following scaling matrix

$$[D_k]_{ii}(\sigma) = (x_i^{(k)} + \gamma) q_i(\sigma). \quad (13)$$

Thus, it is reasonable to determine the steplength parameter according to the quasi-Newton approach

$$\min_{\sigma \in [\sigma_{min}, \sigma_{max}]} \|D_k(\sigma)^{-1} \mathbf{s}^{(k)} - \mathbf{w}^{(k)}\|^2,$$

$$\min_{\sigma \in [\sigma_{min}, \sigma_{max}]} \|\mathbf{s}^{(k)} - D_k(\sigma) \mathbf{w}^{(k)}\|^2,$$

where $\mathbf{s}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}$ and $\mathbf{w}^{(k)} = \nabla f(\mathbf{x}^{(k)}) - \nabla f(\mathbf{x}^{(k-1)})$. The previous one-dimensional minimum problems can be easily solved (for example by means of the `fminbnd` Matlab function), giving two possible values for the steplength σ_k . In our experiments, we adopt the same adaptive alternation strategy applied in [10] for the two Barzilai–Borwein rules.

Following the suggestion in [13], we used $L_1^{(k)} = L_1 = 50$ ($k \in \mathbb{N}$) inner iterations for the image step, $L_2^{(k)} = L_2 = 1$ ($k \in \mathbb{N}$) iteration for the PSF step, a constant image as $\mathbf{x}^{(0)}$ and the autocorrelation of the ideal PSF of LBT as $\boldsymbol{\omega}^{(0)}$. The outer iterations have been arbitrarily stopped at 3000.

In Figure 2 we show the reconstruction of the PSF provided by the three approaches together with the horizontal and vertical central cuts of the pictures compared with those of the target PSF. Moreover, in Figure 3 we plotted the reconstruction errors and the decrease of the objective function versus the number of iterations, where for the PSF we used the standard root mean square error (RMSE) $\|\boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}^*\| / \|\boldsymbol{\omega}^*\|$ while for the image we computed the RMSE for each star $|x_i^{(k)} - x_i^*| / |x_i^*|$ ($i = 1, \dots, 100$) and then calculated the mean of the 100 resulting values. From the results obtained

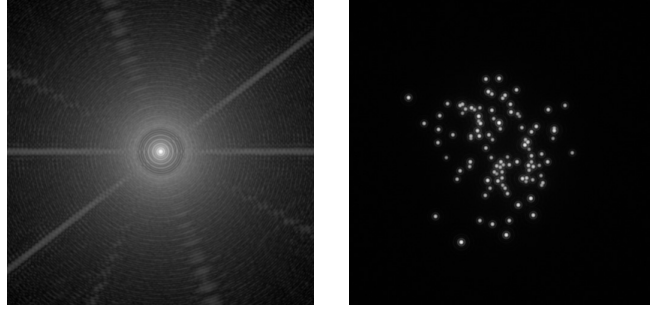


Fig. 1. Star cluster test problem: original PSF (left) and blurred and noisy image (right). Both images are in log scale.

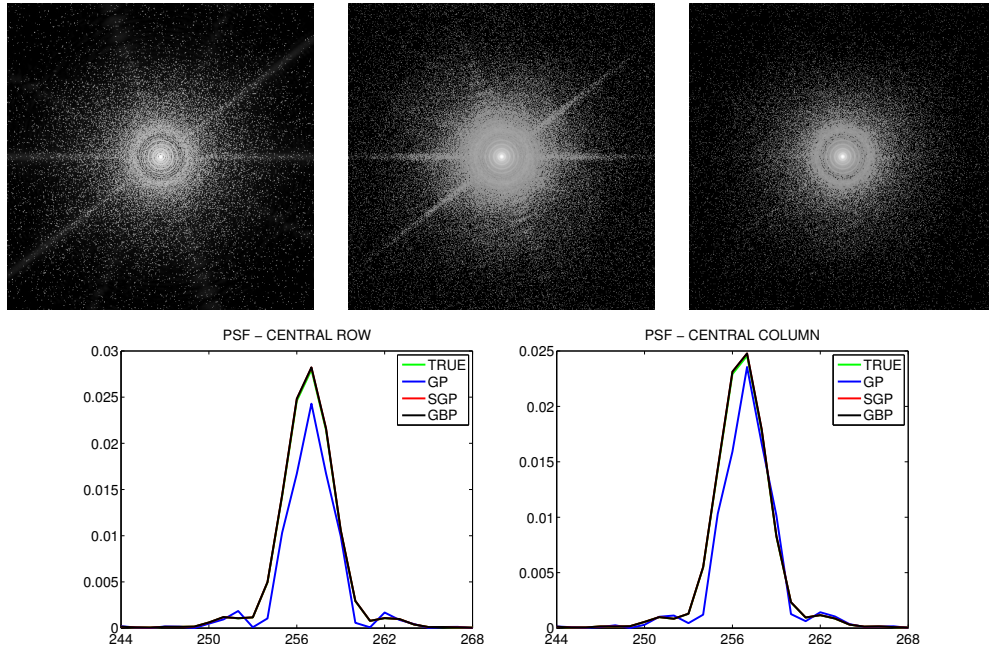


Fig. 2. PSF for the star cluster test problem. First row: reconstructions with GP (left), SGP (middle) and GBP (right) in log scale. Second row: horizontal and vertical central cuts of original and restored PSFs (red and black plots coincide).

in this test problem, we can conclude that the choice of the projection strongly affects the behaviour of the minimization algorithm, and the consequences are emphasized in the case of nonconvex objective function with multiple stationary points. The SGP and GBP choices seem to be attracted by the same limit point, even if going through different paths. With these approaches the reconstructions are very satisfactory, since both the image and the PSF are restored with an error below 1%. On the contrary, the standard projection in Euclidean norm leads to a significantly different pair (x, ω) , with a higher precision in recovering the correct magnitude of the stars coupled with a worse reconstruction of the PSF (RMSE > 20%), as clearly attested also by the plots shown in the second row of Figure 2.

REFERENCES

- [1] D. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, 1999.
- [2] L. Grippo and M. Sciandrone, “On the convergence of the block nonlinear Gauss-Seidel method under convex constraints,” *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.
- [3] S. Bonettini, “Inexact block coordinate descent methods with application to the nonnegative matrix factorization,” *IMA J. Numer. Anal.*, vol. 31, no. 4, pp. 1431–1452, 2011.
- [4] A. Cassioli, D. Di Lorenzo, and M. Sciandrone, “On the convergence of inexact block coordinate descent methods for constrained optimization,” *Eur. J. Oper. Res.*,

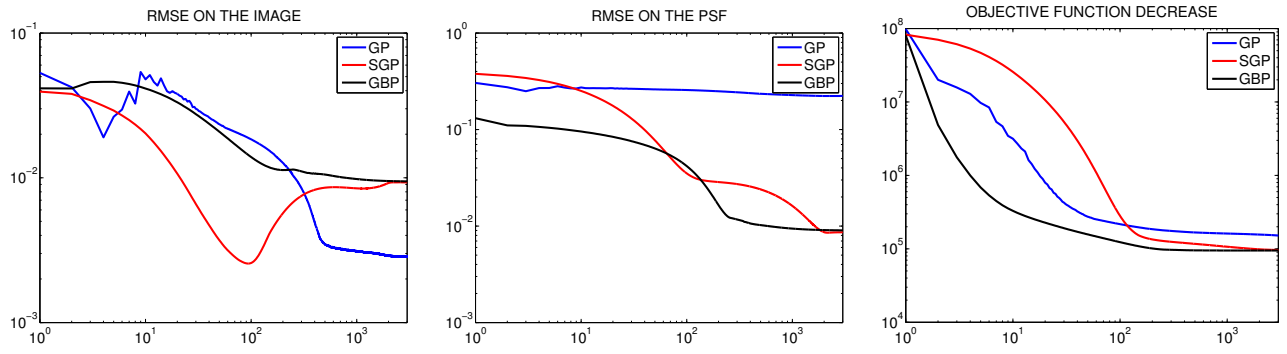


Fig. 3. Blind deconvolution: RMSE on image (left), RMSE on PSF (middle) and objective function decrease (right) in logarithmic scale versus the iterations number (best RMSE on image in unblind case $\approx 1.6 \cdot 10^{-4}$).

vol. 231, no. 2, pp. 274–281, 2013.

- [5] S. Bonettini, M. Prato, and S. Rebegoldi, “A cyclic block coordinate descent method with generalized gradient projections,” *ArXiv e-prints*, p. 1502.06737, 2015.
- [6] J. Bolte, S. Sabach, and M. Teboulle, “Proximal alternating linearized minimization for nonconvex and nonsmooth problems,” *Math. Program.*, vol. 146, no. 1–2, pp. 459–494, 2014.
- [7] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, “A block coordinate variable metric forward-backward algorithm,” <hal-00945918>, Feb. 2014.
- [8] P. Frankel, G. Garrigos, and J. Peypouquet, “Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates,” *J. Optim. Theory Appl.*, vol. 165, no. 3, pp. 874–900, 2015.
- [9] E. G. Birgin, J. M. Martinez, and M. Raydan, “Inexact spectral projected gradient methods on convex sets,” *IMA J. Numer. Anal.*, vol. 23, no. 4, pp. 539–559, 2003.
- [10] S. Bonettini, R. Zanella, and L. Zanni, “A scaled gradient projection method for constrained image deblurring,” *Inverse Probl.*, vol. 25, no. 1, pp. 015002, 2009.
- [11] V. N. Mahajan, “Strehl ratio for primary aberrations in terms of their aberration variance,” *J. Opt. Soc. Am.*, vol. 73, no. 6, pp. 860–861, 1983.
- [12] G. Desiderà and M. Carillet, “Strehl-constrained iterative blind deconvolution for post-adaptive-optics data,” *Astron. Astrophys.*, vol. 507, no. 3, pp. 1759–1762, 2009.
- [13] M. Prato, A. La Camera, S. Bonettini, and M. Bertero, “A convergent blind deconvolution method for post-adaptive-optics astronomical imaging,” *Inverse Probl.*, vol. 29, no. 6, pp. 065017, 2013.
- [14] M. Prato, A. La Camera, S. Bonettini, S. Rebegoldi, M. Bertero, and P. Boccacci, “A blind deconvolution method for ground-based telescopes and Fizeau interferometers,” manuscript submitted for publication.
- [15] S. Bonettini and M. Prato, “Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm,” *Inverse Probl.*, vol. 26, no. 9, pp. 095001, 2010.
- [16] S. Bonettini, A. Cornelio, and M. Prato, “A new semi-blind deconvolution approach for Fourier-based image restoration: an application in astronomy,” *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1736–1757, 2013.
- [17] S. Bonettini and M. Prato, “Accelerated gradient methods for the X-ray imaging of solar flares,” *Inverse Probl.*, vol. 30, no. 5, pp. 055004, 2014.
- [18] I. Loris, M. Bertero, C. De Mol, R. Zanella, and L. Zanni, “Accelerating gradient projection methods for ℓ_1 -constrained signal recovery by steplength selection rules,” *Appl. Computat. Harmon. A.*, vol. 27, no. 2, pp. 247–254, 2009.
- [19] M. Prato, R. Cavicchioli, L. Zanni, P. Boccacci, and M. Bertero, “Efficient deconvolution methods for astronomical imaging: algorithms and IDL-GPU codes,” *Astron. Astrophys.*, vol. 539, pp. A133, 2012.
- [20] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, “Efficient gradient projection methods for edge-preserving removal of Poisson noise,” *Inverse Probl.*, vol. 25, no. 4, pp. 045010, 2009.
- [21] R. Zanella, G. Zanghirati, R. Cavicchioli, L. Zanni, P. Boccacci, M. Bertero, and G. Vicidomini, “Towards real-time image deconvolution: application to confocal and sted microscopy,” *Sci. Rep.*, vol. 3, pp. 2523, 2013.
- [22] A. Nemirovski, “Prox-method with rate of convergence $o(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems,” *SIAM J. Optim.*, vol. 15, no. 1, pp. 229–251, 2004.