The Scaled Gradient Projection Method: An Application to Nonconvex Optimization

M. Prato[1], **A. La Camera**[2], **S. Bonettini**[3], **and M. Bertero**[2]

[1]Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia, Via Campi 213/b, Modena 41125, Italy
[2]Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi
Università di Genova, Via Dodecaneso 35, Genova 16145, Italy
[3]Dipartimento di Matematica e Informatica
Università di Ferrara, Via Saragat 1, Ferrara 44122, Italy

Abstract— The scaled gradient projection (SGP) method is a variable metric forward-backward algorithm designed for constrained differentiable optimization problems, as those obtained by reformulating several signal and image processing problems according to standard statistical approaches. The main SGP features are a variable scaling matrix multiplying the gradient direction at each iteration and an adaptive steplength parameter chosen by generalizing the well-known Barzilai-Borwein rules.

An interesting result is that SGP can be exploited within an alternating minimization approach in order to address optimization problems in which the unknown can be splitted in several blocks, each with a given convex and closed feasible set. Classical examples of applications belonging to this class are the non-negative matrix factorization and the blind deconvolution problems.

In this work we applied this method to the blind deconvolution of multiple images of the same target obtained with different PSFs. In particular, for our experiments we considered the NASA funded Fizeau interferometer LBTI of the Large Binocular Telescope, which is already operating on Mount Graham and has provided the first Fizeau images, demonstrating the possibility of reaching the resolution of a 22.8 m telescope. Due to the Poisson nature of the noise affecting the measured images, the resulting optimization problem consists in the minimization of the sum of several Kullback-Leibler divergences, constrained in suitable feasible sets accounting for the different features to be preserved in the object and the PSFs.

1. INTRODUCTION

Interferometry is a standard way exploited in astronomical imaging to obtain high angular resolution starting from two or more telescopes with smaller diameters. Famous examples of astronomical interferometers are the Very Large Telescope Interferometer (VLTI), the Navy Prototype Optical Interferometer (NPOI) and the Center for High Angular Resolution Astronomy (CHARA) array. A further Fizeau interferometer which recently provided its first images is that located at the Large Binocular Telescope (LBT, Mount Graham, Arizona) and called Large Binocular Telescope Interferometer (LBTI [1]). With its two 8.4-m primary mirrors, separated by a 14.4-m center-to-center distance, and mounted on a common alt-az mount, LBT can be considered the very first Extremely Large Telescope (ELT). The adaptive secondary mirrors provide high-order adaptive optics corrections which produce images with Strehl Ratio (SR), i.e., the ratio of peak diffraction intensity of an aberrated versus perfect waveform (see, e.g., [9]), greater than 0.9. LBTI collects the light from both primary mirrors and combines the two beams on the image plane (Fizeau imaging). In order to obtain a uniform coverage of the frequency plane, several LBTI images at different rotation angles are needed. Each of these measured images will be affected by several sources of noise, whose statistical nature includes both Poisson (due to, e.g., radiation from the object, background, dark current) and Gaussian (read-out-noise (RON) due to the amplifier) components. It has been shown [14] that the latter one can be handled by adding the variance of the RON to the measured images and the background and considering the resulting data as purely Poissonian. As a result of this, a statistical approach to the image reconstruction problem leads to the minimization of the sum of a given number p (equal to the number of images acquired by LBTI) of Kullback-Leibler divergences

$$\min_{\substack{f\in\Omega_f\\ \omega_1,\dots,\omega_p\in\Omega_\omega}} KL(f,\omega_1,\dots,\omega_p) \equiv \sum_{k=1}^{p}\sum_{i=1}^{n^2}(g_k)_i \ln\left(\frac{(g_k)_i}{(\omega_k\otimes f+b_k)_i}\right)+(\omega_k\otimes f+b_k)_i-(g_k)_i, \qquad (1)$$

where:

- f is the $n \times n$ image to be recovered;
- g_k is the k-th $n \times n$ measured image, and b_k the corresponding $n \times n$ background radiation;
- ω_k is the unknown $n \times n$ point spread function (PSF) related to g_k, which has to be estimated during the reconstruction procedure;
- Ω_f and Ω_ω are the feasible sets of f and ω, accounting for physical constraints that object and PSFs have to satisfy.

Since we are assuming that the PSFs are unknown, problem (1) is a so-called blind deconvolution problem [7], which is highly ill-posed due to the infinite solutions available. Moreover, from the optimization point of view the objective function in (1) is nonconvex, thus leading to the presence of multiple stationary points. A possible way to address this kind of problems is to introduce suitable constraints on the unknown to reduce the set of feasible solutions. In particular, in [11, 12] the following feasible sets have been considered

$$\Omega_f = \left\{ f \in \mathbb{R}^{n^2} | f \geq 0, \sum_{i=1}^{n^2} f_i = \frac{1}{p} \sum_{k=1}^{p} \sum_{i=1}^{n^2} (g_k - b_k)_i \right\}, \quad \Omega_\omega = \left\{ \omega \in \mathbb{R}^{n^2} | 0 \leq \omega \leq s, \sum_{i=1}^{n^2} \omega_i = 1 \right\},$$

where the upper bound s is estimated from the knowledge of the SR. Moreover, in [12] an alternating minimization approach for the solution of (1) has been proposed based on the scaled gradient projection (SGP) method [3].

The aim of this paper is to exploit the same optimization approach to clash with a further defect which characterizes interferometric images. As known, the PSF can be described as the product of the Airy function (representing the PSF of a single mirror) by the interferometric term $(1 + \cos\theta)$, where θ depends on the current pixel and the orientation of the baseline (besides the other parameters of the acquisition system)[1]. If we consider a (small) difference in the two optical paths, i.e., there is a residual error when the system corrects for the so-called "piston error", then the interferometric term becomes $(1 + \cos(\theta + \phi))$. This phase error ϕ moves the position of the fringes in the direction of the baseline (and therefore in the direction orthogonal to the fringes themselves) with the result that the PSF is not symmetric with respect to the original PSF. The corresponding image is affected by the same effect.

2. THE CYCLIC BLOCK COORDINATE GRADIENT PROJECTION METHOD

The cyclic block coordinate gradient projection (CBCGP) method is an alternating minimization algorithm originally proposed in [2] in the context of non-negative matrix factorization [6] and adapted in following works to astronomical blind deconvolution problems [4, 11, 12]. The main idea at the basis of this approach is a partial minimization over each block of variables in a Gauss-Seidel style performed inexactly by means of the scaled gradient projection (SGP) method [3]. The SGP algorithm is a first-order method which applies to a general constrained optimization problem

$$\min_{x \in \Omega} \varphi(x). \tag{2}$$

The main steps of SGP are reported in Algorithm 1, where $\mathcal{D}$ is the set of diagonal positive definite matrices with diagonal elements bounded in a positive interval. If φ is equal to the Kullback-Leibler divergence and the PSFs are normalized to unit volume, a frequently used scaling matrix is the one borrowed from the Richardson-Lucy method [8, 13] and defined as

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

with $\mu > 1$ (see, e.g., [3, 5, 10]). As concerns the choice of the steplength parameter α_k, we adopted a suitable alternation of the Barzilai and Borwein rules (see [3, 10] for further details).

The CBCGP scheme applied to problem (1) is reported in Algorithm 2.

[1] For more details: $\theta(n_1, n_2) = A(n_1 \cos\alpha + n_2 \sin\alpha)$, where α is the hour angle of the baseline and $A = 2\pi\Delta B/\lambda$, being Δ the pixel size of the image in rad/px, B the distance between the two mirrors and λ the observed wavelength.

Algorithm 2: Scaled gradient projection (SGP) method

Choose the starting point $x^{(0)} \in \Omega$, set the parameters $\beta, \delta \in (0,1)$, $0 < \alpha_{\min} < \alpha_{\max}$.

For $k = 0, 1, 2, \ldots$ do the following steps:

Step 1. Choose the steplength parameter $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and the scaling matrix $D_k \in \mathcal{D}$;

Step 2. Projection: $y^{(k)} = P_{\Omega, D_k^{-1}}(x^{(k)} - \alpha_k D_k \nabla\varphi(x^{(k)}))$;

Step 3. Descent direction: $d^{(k)} = y^{(k)} - x^{(k)}$;

Step 4. Set $\lambda_k = 1$;

Step 5. Backtracking loop:
If $\varphi(x^{(k)} + \lambda_k d^{(k)}) \leq \varphi(x^{(k)}) + \beta\lambda_k \nabla\varphi(x^{(k)})^T d^{(k)}$ then
go to Step 6;
Else
set $\lambda_k = \delta\lambda_k$ and go to Step 5.
Endif

Step 6. Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$.

End

Algorithm 3: Cyclic block coordinate gradient projection (CBCGP) method

Choose the starting points $f^{(0)} \in \Omega_f$, $\omega_1^{(0)}, \ldots, \omega_p^{(0)} \in \Omega_\omega$, and two integers $N_f, N_\omega \geq 1$.

For $h = 0, 1, 2, \ldots$ do the following steps:

Step f. Choose an integer $1 \leq N_f^{(h)} \leq N_f$ and compute $f^{(h+1)}$ by applying $N_f^{(h)}$ iterations of Algorithm 2 to problem

$$\min_{f \in \Omega_f} KL(f, \omega_1^{(h)}, \ldots, \omega_p^{(h)}) \tag{3}$$

starting from the point $f^{(h)}$.

Step ω. For $k = 1, \ldots, p$ do the following steps:

Choose an integer $1 \leq N_{\omega_k}^{(h)} \leq N_\omega$ and compute $\omega_k^{(h+1)}$ by applying $N_{\omega_k}^{(h)}$ iterations of Algorithm 2 to problem

$$\min_{\omega \in \Omega_\omega} KL(f^{(h+1)}, \omega_1^{(h+1)}, \ldots, \omega_{k-1}^{(h+1)}, \omega, \omega_{k+1}^{(h)}, \ldots, \omega_p^{(h)}) \tag{4}$$

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

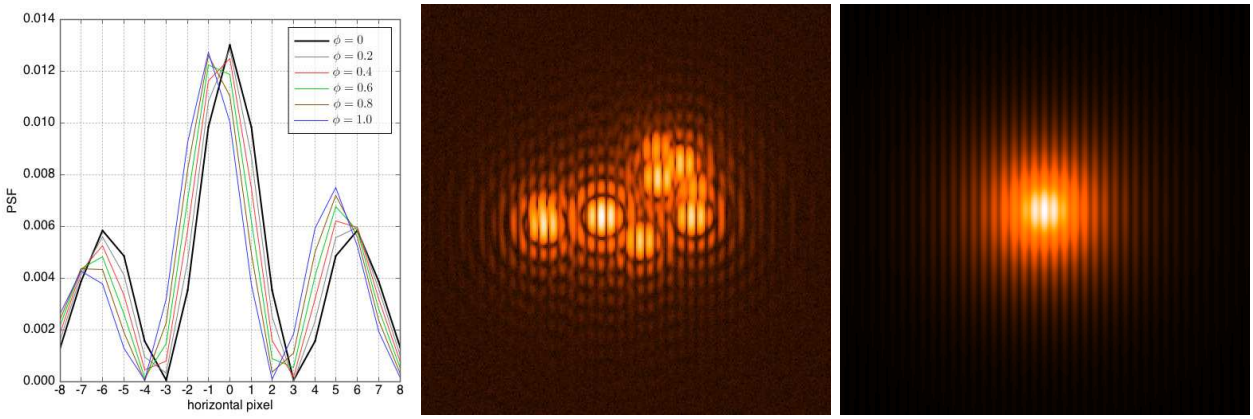
End

End

3. NUMERICAL EXPERIMENTS

We simulated several diffraction limited, SR = 1, LBTI PSFs changing the values of the mentioned phase error. Each set of PSFs is a cube of three PSFs with baseline angles at 0°, 60°, and 120° and corresponds to a different value of ϕ (ranging from 0, 0.2, ..., 1.0). An horizontal cut of the 0° PSFs is shown in the left panel of Fig. 1. By choosing a model of star cluster (based on the nine brightest stars of the Pleiades, as described in [10, 11]), we used these PSFs for generating several multiple images. We adopted a pixel-size of ~ 11 mas and we supposed to acquire images in M-band ($\lambda = 4.8\,\mu$m).

We applied our blind algorithm by using as initialization the autocorrelation of the ideal diffraction limited PSF (i.e., the PSF with no phase error). We remark that this choice has a SR of about 0.35, as described in our previous papers, with positions of the fringes that do not correspond with those in the images. Therefore our goal is to verify whether the algorithm is able to reconstruct both the object and the PSFs and what is the maximum phase error which gives satisfying results. Following [11, 12], we chose $N_f^{(h)} = 50$ and $N_{\omega_k}^{(h)} = 1$ for all h, k, and we performed 1000 outer

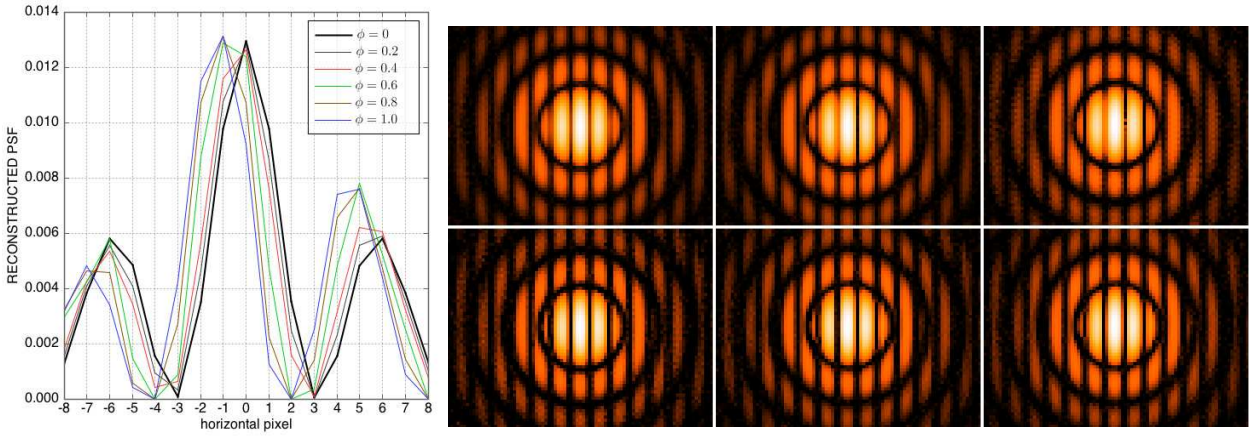


(a) (b) (c)

Figure 1: (a) The horizontal cut of the simulated PSFs for each of the phase error ϕ and for the baseline angle of 0°. A positive value moves the position of the fringes towards left. (b) The input image corresponding to $\phi = 1.0$ for the baseline angle of 0°. (c) The 0° PSF used for the initialization. The images are shown in log scale with a zoom factor of 2.

Table 1: Astrometric error (AE), magnitude average relative error (MARE) and root mean square error (RMSE) for different phase errors ϕ. For the three PSFs, the values in brackets denote the errors after re-centering. The normalized value of the objective function is also shown.

ϕ	AE (pixels)	MARE	RMSE_{0°	RMSE_{60°	RMSE_{120°	$2/(pn^2)KL$
0.0	< 0.01	7.5×10^{-5}	0.7%	0.7%	0.7%	0.7318
0.2	< 0.01	1.0×10^{-4}	0.6%	0.6%	0.6%	0.7363
0.4	0.22	7.6×10^{-5}	7.9% (4.5%)	28.0% (16.5%)	26.6% (16.2%)	0.8417
0.6	0.92	1.1×10^{-4}	35.4% (15.9%)	45.2% (18.3%)	20.0% (11.8%)	0.9081
0.8	1.26	9.6×10^{-5}	37.4% (20.0%)	64.1% (9.3%)	39.1% (9.2%)	0.8234
1.0	1.27	1.1×10^{-4}	37.5% (18.8%)	64.5% (9.1%)	39.0% (9.1%)	0.8192



(a) (b)

Figure 2: (a) The horizontal cut of the reconstructed PSFs for each of the phase error ϕ. (b) The six reconstructed PSFs corresponding to $\phi = 0, 0.2, \ldots, 1.0$ (in order from left to right, from top to bottom) and for the baseline angle of 0°. The images are shown in log scale with a zoom factor of 4.

iterations. In order to evaluate the quality of the reconstructions we computed three errors: the astrometric error (AE), the magnitude average relative error (MARE), and the root mean square

error (RMSE). The three errors are defined as:

$$\mathrm{AE} = \frac{1}{q}\sum_{i=1}^{q}\sqrt{(x_i-\tilde{x}_i)^2+(y_i-\tilde{y}_i)^2}; \quad \mathrm{MARE} = \frac{1}{q}\sum_{i=1}^{q}\frac{|m_i-\tilde{m}_i|}{\tilde{m}_i}; \quad \mathrm{RMSE}_k = \frac{\|\omega_k-\tilde{\omega}_k\|_2}{\|\tilde{\omega}_k\|_2}$$

where q is the number of stars, $(x_i,\, y_i)$ and $(\tilde{x}_i,\, \tilde{y}_i)$ are the reconstructed centroid and the original position of each star, m_i and $\tilde{m}_i$ are the reconstructed and the true magnitudes, ω_k and $\tilde{\omega}_k$ are the reconstructed and true PSFs and $\|\cdot\|_2$ denotes the Euclidean norm.

All the errors are reported in Table 1. As concerns RMSE, it is interesting to note that the values are quite high for all cases with $\phi > 0.2$. Indeed, since each star is reconstructed with a small, but significant, astrometric error, this causes a small displacement of the reconstructed PSF with respect to the original one used in the computation of the RMSE. Therefore we re-centered each PSF by means of a sub-pixel-precision procedure and we computed again the RMSE (the new values are shown in parenthesis). Of course, in case of real images this operation is not necessary.

In conclusion, the proposed strategy seems to provide good reconstructions of the PSFs only if the measured images are affected by a low phase error, although the magnitude of the stars are in general well recovered. Future work will include the addition of explicit regularization terms in the objective function specifically aimed at preserving the features which characterize the LBTI PSFs.

ACKNOWLEDGMENT

The present paper has been supported by the Italian Ministry for University and Research (projects FIRB — Futuro in Ricerca 2012, contract RBFR12M3AC, and PRIN 2012, contract 2012MTE38N) and the Italian Gruppo Nazionale per il Calcolo Scientifico (GNCS).

REFERENCES

1. Bailey, V. P., P. M. Hinz, PA. T. uglisi, et al., "Large binocular telescope interferometer adaptive optics: On-sky performance and lessons learned," *Proc. SPIE*, Vol. 9148, 914803, 2014.
2. Bonettini, S., "Inexact block coordinate descent methods with application to the nonnegative matrix factorization," *IMA J. Numer. Anal.*, Vol. 31, No. 4, 1431–1452, 2011.
3. Bonettini, S., R. Zanella, and L. Zanni, "A scaled gradient projection method for constrained image deblurring," *Inverse Probl.*, Vol. 25, No. 1, 015002, 2009.
4. Bonettini, S., A. Cornelio, and M. Prato, "A new semiblind deconvolution approach for Fourier-based image restoration: An application in astronomy," *SIAM J. Imaging Sci.*, Vol. 6, No. 3, 1736–1757, 2013.
5. Bonettini, S. and M. Prato, "Accelerated gradient methods for the X-ray imaging of solar flares," *Inverse Probl.*, Vol. 30, No. 5, 055004, 2014.
6. Lee, D. D. and H. S. Seung, "Learning the part of objects from non-negative matrix factorization," *Nature*, Vol. 401, No. 6755, 788–791, 1999.
7. Levin, A., Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," *IEEE Conference on Computer Vision and Pattern Recognition, 2009, CVPR 2009*, 1964–1971, Paris, France, Jun. 2009.
8. Lucy, L. B., "An iterative technique for the rectification of observed distributions," *Astronom. J.*, Vol. 79, No. 6, 745–754, 1974.
9. Mahajan, V. N., "Strehl ratio for primary aberrations in terms of their aberration variance," *J. Opt. Soc. Am.*, Vol. 73, No. 6, 860–861, 1983.
10. Prato, M., R. Cavicchioli, L. Zanni, P. Boccacci, and M. Bertero, "Efficient deconvolution methods for astronomical imaging: Algorithms and IDL-GPU codes," *Astron. Astrophys.*, Vol. 539, A133, 2012.
11. Prato, M., 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, 065017, 2013.
12. Prato, M., A. La Camera, S. Bonettini, S. Rebegoldi, M. Bertero, and P. Boccacci, "A blind deconvolution method for ground-based telescopes," *New Astron.*, Vol. 40, 1–13, 2015.
13. Richardson, W. H., "Bayesian based iterative method of image restoration," *J. Opt. Soc. Amer.*, Vol. 62, No. 1, 55–59, 1972.
14. Snyder, D. L., A. M. Hammoud, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Am. A*, Vol. 10, No. 5, 1014–1023, 1993.