A fast subgradient algorithm in image super-resolution

IOP Conf. Series: Journal of Physics: Conf. Series **904** (2017) 012009 doi:10.1088/1742-6596/904/1/012009

A fast subgradient algorithm in image super-resolution

D. Lazzaro[1], E. Loli Piccolomini[1], V. Ruggiero[2], F. Zama[1]

[1] Department of Mathematics, University of Bologna
[2] Department of Mathematics and Computer Science, University of Ferrara

E-mail: fabiana.zama@unibo.it

Abstract. In this paper we propose an $\epsilon-$subgradient method for solving a constrained minimization problem arising in super-resolution imaging applications. The method, compared to the state-of-the-art methods for single image super-resolution on some test problems, proves to be very efficient, both for the reconstruction quality and the computational time.

1. Introduction

Image super-resolution reconstruction is the process of obtaining High Resolution (HR) images from observed Low Resolution (LR) images. The problem of super-resolution is of great importance in many applications, such as medicine or object recognition (face, bar codes, ...), where the electronic imaging devices are equipped with low resolution cameras, while HR images are finally desired. The super-resolution process can be performed from a single image (SISR) or from multiple images of the same scene (MISR), such as in the case of videos or medical imaging devices for example. In this paper, we will consider only the case of SISR.

The SIRS problem is ill-posed, since identical LR images can be generated from different HR images. Furthermore, in addition to being connected by a down-sampling operator, the HR and LR images are related by a Point Spread Function (as a simple convolution, for example), that is an ill-posed operator. The different approaches, present in literature for image super-resolution, can generally be grouped into four categories: interpolation-based algorithms, example-based algorithms, sparse-representation-based algorithms and reconstruction-based algorithms (see [5] and the references therein). The reconstruction-based algorithms are aimed at solving a minimization problem, where the objective function is the sum of a fidelity term related to the data noise and a regularization term, representing the prior on the solution and, at the same time, enabling to face the ill-posedness of the problem. This formulation has the advantage of simultaneously exploiting the degradation model and the information contained in the prior, thus reducing noise and artifacts in the HR reconstructed image [5, 12].

We consider a reconstruction-based algorithm to address a minimization problem where the fidelity term is the ℓ_2-norm, the regularization term is the discrete Total Variation (TV) function and the solution is constrained to be nonnegative. This model has been widely treated for the deblurring applications and many algorithms have been proposed for its solution, such as, for example, [11, 13, 14]. The Bregman method proposed in [14] has been later used in [12] for image super-resolution. In this paper we propose a scaling ϵ-subgradient method presented in [2] for deblurring of images corrupted by Poisson noise. The method has a fast computational

cost, since it doesn't require any linear system solution. We test the algorithm on two problems related to synthetic and natural images and compare the results with some state-of-the-art methods for super-resolution.

The paper is organized as follows. In section 2 we present the mathematical problem formulation; in section 3 we describe the algorithm and finally in section 4 we report the results of some numerical experiments and the final conclusions.

2. Numerical model and notation

Let $G \in \mathbb{R}^{M\times N}$ be a low resolution observed image and let $X \in \mathbb{R}^{(s\cdot M)\times(s\cdot N)}$ be the high resolution image to be recovered, with up-sampling factor $s > 1$. The mathematical model of a LR image formation can be written as:

$$g = SHx + \eta \tag{1}$$

where $g \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ are obtained after vector reordering of the matrices G and X respectively, hence $m = M \cdot N$ and $n = s^2 \cdot M \cdot N$. The matrix H represents the discrete blurring operator acting on the HR data x and S is the down-sampling operator mapping HR blurred data into the LR data. The vector $\eta \in \mathbb{R}^m$ represents additive data noise.

The numerical model proposed in the reconstruction-based super-resolution algorithms is given by the following constrained minimization problem:

$$min_{x\in\mathbb{R}^n} \|SHx - g\|_2^2 + \beta TV(x) \quad s.t.\ x \geq 0 \tag{2}$$

where the nonnegativity constraint on the solution is a physical requirement on the components of the image to be recovered. Here $\beta > 0$ is the regularization parameter and the discrete TV can be written as

$$TV(x) = \sum_{i=1}^{n} \|A_i x\|$$

where $A_i \in R^{2\times n}$ is the forward difference approximation of the gradient of x at the pixel i.
Under the usual assumption on the blurring operator, such as

$$H_{i,j} \ \geq \ 0 \ \ ; \ \ \sum_{i=1}^{n} H_{i,j} \ > \ 0 \ \ , \ \ \forall j = 1, ..., n \ \ ; \ \ \sum_{j=1}^{n} H_{i,j} \ > \ 0 \ \ , \ \ \forall i = 1, ..., n, \tag{3}$$

the problem (2) is coercive and then a solution exists. Let's denote $\mathbf{1}$ as the vector with all entries equal to 1. Since $H\mathbf{1} \neq \mathbf{0}$ and $\mathcal{N}(SH) = \{x : (Hx)_i = 0 \text{ for } i \in I\}$, where $I \neq \emptyset$ is the set of the non zero columns of S, the intersection of the space of constant images $\alpha\mathbf{1}$ (which is the null space of $TV(x)$) with the null space of SH is just the zero vector.
In the following, we denote by $A \in \mathbb{R}^{2n\times n}$ the block matrix $(A_1^T, A_2^T, ...A_n^T)^T$. Furthermore, given a symmetric positive definite matrix D, $P_{D,C}(x)$ denotes the projection on the closed set C with respect to the metric induced by D, defined as $P_{D,C}(x) = \text{argmin}_{z\in C} \|z - x\|_D^2$, where $\|y\|_D^2 = y^T Dy$. For $D = I$, we have the standard Euclidean projection on the set C. The standard Euclidean norm is denoted by $\|\cdot\|$.

3. The scaled subgradient algorithm

The considered application is a special case of the problem

$$\min_{x\in\mathbb{R}^n} f_0(x) + f_1(Ax) + \Phi(x), \tag{4}$$

where $f_0(x) = \frac{1}{2}\|SHx - g\|^2$ is the discrepancy function, $f_1(Ax) = \beta TV(x)$ is the discrete Total Variation and $\Phi(x)$ is the indicator function $\iota_X(x)$ of the non-negative orthant X. Using the Fenchel transform of $f_1(Ax)$, which is the indicator function $i_Y(y)$ of the bounded and convex set $Y = \{(y_1^T, ...y_n^T)^T | y_i \in \mathbb{R}^2,\ \|y_i\| \leq 1\}$, from the equality $f_1(Ax) = \beta \max_{y\in Y}(Ax)^T y$, we obtain the equivalent primal–dual formulation of the problem (2) and we can address the numerical solution by the class of Primal–Dual Hybrid Gradient (PDHG) methods (see [17, 4] and references therein). Recently in [2] a variable metric has been introduced in a PDHG method and, by interpreting this scheme as an ϵ-subgradient iteration, a dynamic computation for the primal stepsize has been proposed. For the problem (2), the basic iteration of the Scaled Primal–Dual Hybrid Gradient (SPDHG) method requires only matrix–vector operations and simple projections:

$$y^{(k+1)} = P_{I,Y}(y^{(k)} + \tau_k \beta A x^{(k)}) \tag{5}$$

$$u^{(k)} = d^{(k)} + \beta A^T y^{(k+1)} \tag{6}$$

$$x^{(k+1)} = P_{D_k^{-1},X}\left(x^{(k)} - \frac{\alpha_k}{\min(1, \|u^{(k)}\|_{D_k})} D_k u^{(k)}\right) \tag{7}$$

where $d^{(k)} = H^T S^T (SHx^{(k)} - g)$, $\{D_k\}$ is a sequence of symmetric positive definite matrices with bounded eigenvalues and $\{\tau_k\}$, $\{\alpha_k\}$ are the dual and primal steplength sequences respectively. Method (5)–(7) is a special case of a scaled forward-backward ϵ-subgradient method, where the forward step (5)–(6) provides an ϵ-subgradient of $f_0 + f_1 \circ A$ and the backward or proximal step (7) of $i_X(x)$ is the projection on the non-negative orthant with respect to the variable metric induced by D_k^{-1}. The key point of the interpretation of the forward step is based on the additivity of the ϵ-subgradient and on the fact that the vector $y^{(k+1)}$ defined in (5) belongs to the domain of the Fenchel transform of f_1. Consequently, $\beta A^T y^{(k+1)}$ is an ϵ_k-subgradient of $f_1 \circ A$ at $x^{(k)}$ with $\epsilon_k = f_1(Ax^{(k)}) + \iota_Y(y^{(k+1)}) - \beta {y^{(k+1)}}^T A x^{(k)}$. Furthermore, $\epsilon_k \leq \frac{\mathrm{diam}(Y)^2}{2\tau_k}$ (see [3, Lemma 1]) and, for an *a-priori* selected divergent sequence $\{\tau_k\}$, $\epsilon_k \to 0$, as $k \to \infty$.
In the backward step we use an adaptive stepsize selection strategy for the sequence $\{\alpha_k\}$, inspired by the idea of *level algorithm* of [9]. The resulting algorithm, detailed in Algorithm 1, is the version of SSL algorithm described in [2] for the problem (2). In particular, in the scheme we have $f_k^{rec} = \min_{i=0,...,k}(f_0(x^{(i)}) + f_1(Ax^{(i)}))$, while l is the number of times that the value f^{lev} has been updated and $k(l)$ is the iteration where the l-th updating occurred. Finally, σ_k is the cumulative path length between two successive updates of f^{lev}. Steps 2-5 aim to provide in f_k^{lev} an estimate of the optimal function value at the iterate k, which is used as target level for the successive iterates until the objective function value is sufficiently close to it or the iterates move through a long path without approaching it. In the first case, i.e. when the inequality at Step 3 is satisfied, f_k^{lev} is reduced at Step 5 by subtracting the positive quantity δ_l to the best value obtained so far, f^{rec}. In the other case, when the inequality at Step 4 is satisfied, the estimated difference from the optimal value δ_l is reduced and, as a consequence of Step 5, the target level f_k^{lev} is increased. Assuming to choose an a-priori dual stepsize τ_k such that $\lim_{k\to\infty} \tau_k = \infty$ and a sequence of positive definite scaling matrices $\{D_k\}$ converging to the identity matrix as $k \to \infty$, the convergence of the SPDHG method with adaptive updating rule for α_k is assured by Corollary 5.2 in [2]. For completeness, we report the convergence proposition for the problem (2).
Proposition. Let $\{x^{(k)}\}$ be the sequence generated by Algorithm 1, where $u^{(k)}$ in (7) is given by (6) with $d^{(k)} = H^T S^T (SHx^{(k)} - g)$. Assume that that there exists $\rho > 0$ such that $\|d^{(k)}\| \leq \rho$ for all k. Define $L_k = \max(\|D_k\|, \|D_k^{-1}\|)$ and assume that $\lim_{k\to\infty} \tau_k = \infty$, $L_k \leq \sqrt{1+\gamma_k}$, $\gamma_k = \mathcal{O}\left(\frac{1}{k^q}\right)$ with $q > 1$. Then we have $\liminf_{k\to\infty} f_0(x^{(k)}) + f_1(Ax^{(k)}) = \min_{x\in X}(f_0(x) + f_1(Ax))$.

Proof. Since $d^{(k)} = \nabla f_0(x^{(k)})$, by Lemma 1 in [3] we have $u^{(k)} \in \partial_{\epsilon_k}(f_0 + f_1 \circ A)(x^{(k)})$, where $\epsilon_k = f_1(Ax^{(k)}) + \iota_Y(y^{(k+1)}) - \beta {y^{(k+1)}}^T Ax^{(k)}$, with $\epsilon_k \leq \frac{\mathrm{diam}(Y)^2}{2\tau_k}$. Since $\lim_{k\to\infty} \tau_k = \infty$, we have $\lim_{k\to\infty} \epsilon_k = 0$. Choosing a sequence of null subgradients at $x^{(k)}$ for i_X, by Theorem 4.1 in [2], we obtain the result.

Algorithm 1 Scaled Primal–Dual Hybrid Gradient (SPDHG) method for problem (2)

Choose $B > 0$, $\nu_1, \nu_2 \in (0,1)$, $f_{-1}^{rec} = \infty$; $k = 0$, $l = 0$, $k(l) = 0$, $\delta_0 > 0$; choose $x^{(0)} \in X$.

FOR $k = 0, 1, 2, ...$

STEP 0. Computation of ϵ-subgradient $u^{(k)}$ of $f_0(x^{(k)}) + f_1(Ax^{(k)})$ with (5)-(6)
STEP 1. Computation of $f_0(x^{(k)}) + f_1(Ax^{(k)})$
STEP 2. If $f_0(x^{(k)}) + f_1(Ax^{(k)}) < f_{k-1}^{rec}$, then $f_k^{rec} = f_0(x^{(k)}) + f_1(Ax^{(k)})$ else $f_k^{rec} = f_{k-1}^{rec}$.
STEP 3. If $f_0(x^{(k)}) + f_1(Ax^{(k)}) < f_{k(l)}^{rec} - \nu_1\delta_l$, then $k(l+1) = k$, $\sigma_k = 0$, $\delta_{l+1} = \delta_l$, $l = l+1$ and go to Step 5.
STEP 4. If $\sigma_k > B$, then $k(l+1) = k$, $\sigma_k = 0$, $\delta_{l+1} = \nu_2\delta_l$, $l = l+1$.
STEP 5. Set $f_k^{lev} = f_{k(l)}^{rec} - \delta_l$.
STEP 6. Update the stepsize and compute the new iterate

$$\alpha_k = \frac{f_0(x^{(k)}) + f_1(Ax^{(k)}) - f_k^{lev}}{\max(1, \|u^{(k)}\|_{D_k})},$$

$$x^{(k+1)} = P_{D_k^{-1},X}\left(x^{(k)} - \alpha_k D_k \frac{u^{(k)}}{\max(1, \|u^{(k)}\|_{D_k})}\right), \tag{8}$$

STEP 7. $\sigma_{k+1} = \sigma_k + \alpha_k$ and go to Step 1.

END

In order to fully define the SPDHG method, we focus on the strategy to compute a suitable scaling matrix D_k, adapting to our case the split gradient strategy proposed in [1] for nonnegatively constrained differentiable problems, which demonstrated to be very effective in several applications (see [16, 6] and references therein). The key point of this approach consists in finding a subgradient decomposition of the form $u^{(k)} = V(x^{(k)}) - U(x^{(k)})$ with $V(x^{(k)}) > 0$ and $U(x^{(k)}) \geq 0$ for all k and then defining D_k in (7) as a diagonal scaling matrix whose entries are the projection of $x_i^{(k)}/V_i(x^{(k)})$ onto the set $[1/\sqrt{1+\gamma_k}, \sqrt{1+\gamma_k}]$. Thus, we have to find a decomposition of the vector $u^{(k)} = \nabla f_0(x^{(k)}) + \beta A^T y^{(k+1)}$ as the difference of two nonnegative terms and we put $V(x^{(k)}) = H^T S^T SHx^{(k)} + V_R(x^{(k)})$. Indeed, for the first term of $u^{(k)}$, we observe that $H^T S^T SHx \geq 0$ for all $x \geq 0$, while the nonnegative term $V_R(x)$ of the decomposition of the vector $\beta A^T y^{(k+1)}$ can be obtained by a non expensive recurrence formula detailed in [2].

4. Numerical experiments

In this section we report some preliminary results obtained by SPDHG method applied to the problem of image super-resolution. Our aim is to evaluate its best performance in terms of reconstruction accuracy and efficiency and compare our results with those obtained by a few representative state-of-the-art algorithms for image super-resolution. To this purpose we select four representative SISR approaches such as Bi-cubic Interpolation, TIP-ASDS-IR [7], ELAD-TIP14 [15], NCSR [8] and perform the image reconstructions under the same testing conditions

used for SPDHG method. The source codes of TIP-ASDS-IR, ELAD-TIP14 and NCSR methods were downloaded from their relative homepages while the Bi-cubic Interpolation method is realized by means of the Matlab function `imresize()`.
We focus here on two test images relative to synthetic and natural images respectively. The Low Resolution (LR) test images are obtained by High Resolution (HR) images (256×256 pixels), normalized in the range [0,1],(see figure 1 (a)) . The HR images are blurred by a 7×7 Gaussian kernel with standard deviation 1.6 and then down-sampled by a scaling factor equal to 3 both in the horizontal and vertical directions. All the numerical computations are performed using MATLAB R2016b on an Intel i7-3770 CPU with 16 GB RAM.
The reconstruction accuracy is measured using the Peak Signal to Noise Ratio parameter (PSNR (dB)) which is reported in table 1 together with the execution times.

We first present the results related to noiseless LR images. The Bi-cubic method is very fast but always obtains worse quality results, therefore it is reported only to appreciate the improvement obtained by the other more complex methods. We observe that, in the case of synthetic images, SPDHG method reaches the best results both in accuracy and execution times (first row of table 1). In the case of natural images (second row of table 1), SPDHG has better quality and execution times compared to TIP-ASDS-IR and ELAD-TIP14 while it has lower PSNR than NCSR. However NCSR results to be the most expensive from the computational point of view. The reconstructed HR images are shown in figure 1.We do not show the images for TIP-ASDS-IR since the PSNR are very close to ELAD-TIP14 while the computation times are larger.

We consider now the case of noisy data, obtained by adding Gaussian white noise with variance $\sigma = 0.005$. We report in table 2 the results obtained with ELAD-TIP14, NCSR and SPDHG, that gave the best results on noiseless data. Figure 2 shows a detail of each reconstructed image. We highlight that the proposed SPDHG method gets the best PSNR values in the shortest time.

We can conclude that SPDHG is a very promising method for fast and accurate super-resolution reconstructions. It allows us to obtain PSNR values better or comparable to those obtained by the best state-of-the-art methods and requires always the lowest computational times.

Image	Bicubic	TIP-ASDS-IR	ELAD-TIP14	NCSR	SPDHG
Synthetic	25.90	30.27 (108.20)	30.48 (8.67)	31.15 (154.29)	**31.8 (3.01)**
Satellite	23.02	25.50 (114.96)	25.42 (8.59)	**26.03** (164.80)	25.7 (**8.18**)

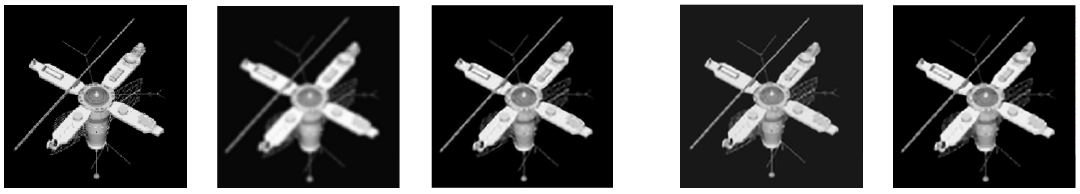
Table 1: PSNR and execution times (in seconds) for each method: best results in bold.

Image	ELAD-TIP14	NCSR	SPDHG
Synthetic	30.00 (8.9)	29.26 (140)	**31.4 (1.8)**
Satellite	25.28 (8.54)	25.60 (149)	**25.61 (4.72)**

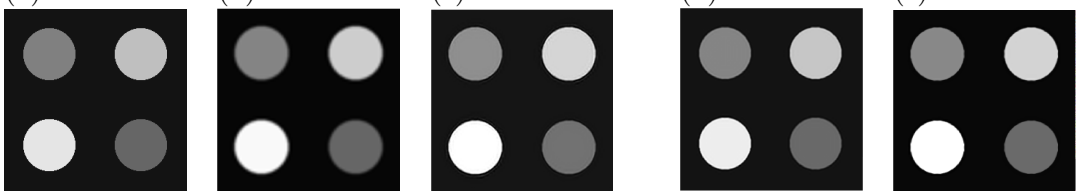
Table 2: PSNR and execution times (in seconds) for the best methods on noisy data ($\sigma = 5{\cdot}10^{-3}$): best results in bold.

Acknowledgments
This work has been partially supported by the Italian Institute GNCS - INdAM.

Figure 1: Original and reconstructed HR Images



(a) Exact (b) Bicubic (c) ELAD-TIP14 (d) NCSR (e) SPDHG



(a) Exact (b) Bicubic (c) ELAD-TIP14 (d) NCSR (e) SPDHG

Figure 2: Reconstructed HR image details: noisy case.

(a) ELAD-TIP14 (b) NCSR (c) SPDHG

(a) ELAD-TIP14 (b) NCSR (c) SPDHG

References

[1] Bertero M., Lanteri H. and Zanni L. 2008 *in Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)* 3763.
[2] Bonettini S., Benfenati A., and Ruggiero V. 2016 *SIAM Journal on Optim.* **26:3** 1741-1772
[3] Bonettini S. and Ruggiero V. 2012 *JMIV* **44**, 236-253
[4] Chambolle A. and Pock T. 2010 *JMIV* **40** 120-145.
[5] Chen H., He X., Teng Q. and Ren C. 2016 *Sig. Proc.: Image Commun.* **43** 68-81.
[6] Coli V. L., Ruggiero V. , Zanni L. 2016 *AIP Conference Proceedings* ,**1776** 040002
[7] Dong W. , Zhang L., Shi G., and Wu X. 2011 *IEEE Transactions on Image Processing*, **20(7)** 1838–1857.
[8] Dong W. , Zhang L., Shi G., and Li X. 2013 *IEEE Transactions on Image Processing* **22(4)** 1620–1630.
[9] Goffin J.L. and Kiwiel K.C. 1999 *Mathematical Programming* **85** 207-211.
[10] Jiang J., Ma X. , Chen C., Lu T., Wang Z., and Ma J. 2017 *IEEE Transaction on Multimedia* **19(1)** 15–26.
[11] Landi G. , Loli Piccolomini E. 20134 *Num. Alg.* **62** 487-504
[12] Marquina A., Osher S. 2008 *J. Sci. Comp.* **37** 367-382.
[13] Montefusco L. and Lazzaro D. 2012 *IEEE Transactions on Image Processing* **21(4)** 1676-1686.
[14] Osher S., Burger M., Goldfarb D., Xu J. and Jin W. 2005 *Mult. Mod. Simul.***4** 460-489.
[15] Peleg T. and Elad M. 2014 *IEEE Transactions on Image Processing* **23(6)** 2569–2582.
[16] Zanella R., Boccacci P., Zanni L., Bertero M. 2009 *Inverse Probl.* **25** 045010.
[17] Zhu M. and Chan T., 2008 *UCLA CAM Report* [08-34]