GRADIENT PROJECTION APPROACHES FOR OPTIMIZATION PROBLEMS IN IMAGE DEBLURRING AND DENOISING

Silvia Bonettini[1], *Federico Benvenuto*[2], *Riccardo Zanella*[3], *Luca Zanni*[3], *and Mario Bertero*[4]

[1]Department of Mathematics, University of Ferrara,
Scientific-Technological Campus, Building B, Via Saragat, 1, I-44100 Ferrara, Italy
phone: + (39) 0532 974785, fax: + (39) 0532 974787 , email: silvia.bonettini@unife.it

[2]Department of Mathematics, University of Genova,
[4]Department of Computer and Information Sciences, University of Genova,
Via Dodecaneso, 35, I-16146 Genova, Italy
phone: + (39) 010 3536733, fax: + (39) 010 3536699, email: benvenut@dima.unige.it, bertero@disi.unige.it

[3]Department of Mathematics, University of Modena and Reggio Emilia,
Via Campi, 213/B, I-41100 Modena, Italy
phone: + (39) 059 2055206, fax: + (39) 059 370513, email: riccardo.zanella@unimore.it, luca.zanni@unimore.it

ABSTRACT

Gradient type methods are widely used approaches for nonlinear programming in image processing, due to their simplicity, low memory requirement and ability to provide medium-accurate solutions without excessive computational costs. In this work we discuss some improved gradient projection methods for constrained optimization problems in image deblurring and denoising. Crucial feature of these approaches is the combination of special steplength rules and scaled gradient directions, appropriately designed to achieve a better convergence rate. Convergence results are given by exploiting monotone or nonmonotone line-search strategies along the feasible direction. The effectiveness of the algorithms is evaluated on the problems arising from the maximum likelihood approach to the deconvolution of images and from the edge-preserving removal of Poisson noise. Numerical results obtained by facing large scale problems involving images of several mega-pixels on graphics processors are also reported.

1. INTRODUCTION

Image deblurring and denoising procedures based on the maximum likelihood criterion lead to model the imaging problem as a constrained large scale optimization problem, with a nonlinear, convex objective function. In general, given the observed blurred and noisy image $y \in \mathbb{R}^m$, the objective function is a combination of a fit–to–data functional J_y that measures the discrepancy of the solution from y, and a penalty term J^R governing some properties of the reconstructed image. Thus, an approximation of the original image can be obtained by computing a vector $x \in \mathbb{R}^n$ that solves the minimization problem

$$\min_{x \geq 0} \quad J_y(x) + \tau J^R(x),$$

where τ is a regularization parameter. The fit–to–data term $J_y(x)$ should be chosen according to the noise statistics: if $A \in \mathbb{R}^{m \times n}$ represents the impulse response of the image acquisition system (A equal to the identity matrix in the case of denoising problems), when Gaussian noise is assumed, then the least square functional is employed (see Section 3.1), while the Kullback-Leibler divergence is more appropriate for the Poisson noise (see Section 4). For the regularization term $J^R(x)$, several choices can be considered; we mention the Tikhonov regularization, the sparsity inducing ℓ_1-penalization and the edge preserving Total Variation regularization.

The optimization problems arising in the imaging framework have very special features, which should be taken into account when designing the optimization methods for computing the solution. For example, the sizes of the problems are very large (since images with several mega-pixels need to be processed), the constraints have a very simple form and, often, some of the data cannot be represented in explicit matrix format but they are available only as operators. The optimization methods successfully applied to these problems consist in iterative schemes that can be grouped in two main classes: the explicit methods that use only matrix-vector products at each iteration and the implicit ones that require to solve a linear system per iteration. The former are widely used for their simplicity, low memory requirement and low cost per iteration, while the latter are appealing for their fast convergence. However, the recent advances on the acceleration techniques for gradient-type methods make some explicit approaches well suited to achieve moderate-accurate solutions with very competitive convergence rate.

This work deals with the explicit method proposed in [4] for image deblurring problems and able to face minimization problem of the general form

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

where $\Omega \subset \mathbb{R}^n$ is a closed convex set and $f : \Omega \to \mathbb{R}$ is a continuously differentiable function. The method is a Scaled Gradient Projection (SGP) algorithm based on special scaling techniques and steplength selection rules. As in other popular algorithms for image restoration, such as the Expectation Maximization (EM) method [12] and the Image Space Reconstruction Algorithm (ISRA) [6], the descent directions exploited by SGP are derived by diagonal scalings of $\nabla f(x)$. The selection of the steplength along these directions is per-

formed by taking into account recent ideas on the alternation of the well known Barzilai-Borwein steplength rules [2]. The above strategies are crucial to improve the convergence rate but they are not sufficient to ensure the convergence of the scheme; to this end, line-search strategies along the feasible direction are exploited to ensure a decrease of the objective function (in a monotone or nonmonotone way) during the iterative procedure. SGP has been evaluated in [4] as iterative regularization method for the deconvolution of images corrupted by Poisson noise, that is for the minimization of the Kullback-Leibler function subject to nonnegativity constraints; it has shown very good convergence rate and significant improvements in comparison to the EM method and the Modified Residual Norm Steepest Descent algorithm [1].
Here, we discuss the application of SGP to other imaging problems: the deconvolution of images corrupted by Gaussian noise and the edge-preserving removal of Poisson noise. Furthermore, we present some numerical results on large scale imaging problems obtained by implementing SGP on Graphics Processing Units.
The paper is outlined as follows. In Section 2 we briefly recall the SGP scheme, in Section 3 we describe its behaviour on the considered imaging problems and in Section 4 we evaluate the implementation of SGP on graphics processors.

2. THE SGP ALGORITHM

To recall the SGP scheme for the solution of the problem (1) we need some basic definitions and notations. We denote by $\|\cdot\|_D$ the vector norm induced by the symmetric positive definite matrix $D \in \mathbb{R}^{n\times n}$: $\|x\|_D = \sqrt{x^T D x}$. As usual, the 2-norm of vectors is denoted by $\|\cdot\|$. Given a closed convex set $\Omega \subset \mathbb{R}^n$, we define the projection operator associated to D as

$$P_{\Omega,D}(x) = \text{argmin}_{y\in\Omega}\|y-x\|_D.$$

In the special case in which $D = I$, I being the identity matrix, and Ω is the nonnegative orthant, the projector operator is denoted by $P_+(\cdot)$. Furthermore, let $\mathscr{D}$ be the set of the $n\times n$ symmetric positive definite matrices D such that $\|D\| \le L$ and $\|D^{-1}\| \le L$, for a given threshold $L > 1$.
The main SGP steps are given in Algorithm 1.
At each SGP iteration the vector

$$y^{(k)} = P_{\Omega,D_k^{-1}}(x^{(k)} - \alpha_k D_k \nabla f(x^{(k)}))$$

is defined by combining a scaled steepest descent direction with a projection on Ω. Since the projection is performed by using the projection operator associated to the inverse of the scaling matrix, it is possible to prove that the resulting search direction $\Delta x^{(k)} = y^{(k)} - x^{(k)}$ is a descent direction for the problem (1), that is ${\Delta x^{(k)}}^T \nabla f(x^{(k)}) < 0$. The global convergence of the algorithm is obtained by means of the nonmonotone line–search procedure described in the Step 5, that implies $f(x^{(k+1)})$ lower than the reference value f_{max}. We observe that this line–search reduces to the standard monotone Armijo rule when $M = 1$ ($f_{max} = f(x^{(k)})$). The following convergence result for SGP has been proved in [4] by employing standard techniques (see also [3]).

THEOREM 2.1 *Let $\{x^{(k)}\}$ be the sequence generated by applying the SGP algorithm to the problem* (1) *and assume that the level set $\Omega_0 = \{x \in \Omega : f(x) \le f(x^{(0)})\}$ is bounded. Then, every accumulation point $x^* \in \Omega$ of the sequence $\{x^{(k)}\}$ is a constrained stationary point, that is*

$$\nabla f(x^*)^T (x - x^*) \ge 0, \quad \forall x \in \Omega.$$

Algorithm 1 Scaled Gradient Projection (SGP) Method

Choose the starting point $x^{(0)} \in \Omega$, set the parameters $\beta, \theta \in (0,1)$, $0 < \alpha_{min} < \alpha_{max}$ and fix a positive integer M.

FOR $k = 0, 1, 2, \ldots$ DO THE FOLLOWING STEPS:

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

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

STEP 3. Descent direction: $\Delta x^{(k)} = y^{(k)} - x^{(k)}$;

STEP 4. Set $\lambda_k = 1$ and $f_{max} = \max_{0\le j\le \min(k,M-1)} f(x^{(k-j)})$;

STEP 5. Backtracking loop:
  let $f_{new} = f(x^{(k)} + \lambda_k \Delta x^{(k)})$;
  IF $f_{new} \le f_{max} + \beta\lambda_k \nabla f(x^{(k)})^T \Delta x^{(k)}$ THEN
    go to Step 6;
  ELSE
    set $\lambda_k = \theta\lambda_k$ and go to Step 5.
  ENDIF

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

END

It is worth to stress that any choice of the steplength $\alpha_k \in [\alpha_{min}, \alpha_{max}]$ and of the scaling matrix $D_k \in \mathscr{D}$ are allowed; then, this freedom of choice can be fruitfully exploited for introducing performance improvements.
An effective selection strategy for the steplength parameter is obtained by adapting to the context of the scaling gradient methods the Barzilai-Borwein (BB) rules [2], widely used in standard non-scaled gradient methods. When the scaled direction $D_k \nabla f(x^{(k)})$ is exploited within a step of the form $(x^{(k)} - \alpha_k D_k \nabla f(x^{(k)}))$, the BB steplength rules become

$$\alpha_k^{(1)} = \frac{{r^{(k-1)}}^T D_k^{-1} D_k^{-1} r^{(k-1)}}{{r^{(k-1)}}^T D_k^{-1} z^{(k-1)}}, \quad \alpha_k^{(2)} = \frac{{r^{(k-1)}}^T D_k z^{(k-1)}}{{z^{(k-1)}}^T D_k D_k z^{(k-1)}},$$

where $r^{(k-1)} = x^{(k)} - x^{(k-1)}$ and $z^{(k-1)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)})$. It may be observed that these formulas reduce to the standard BB rules in the case of non-scaled gradient methods, that is when $D_k = I$. The recent literature on the steplength selection in gradient methods suggests to design steplength updating strategies by alternating the two BB rules. We recall the adaptive alternation strategy proposed in [4, 9], that has given remarkable convergence rate improvements in many different applications. Given an initial value α_0, the steplengths α_k, $k = 1, 2, \ldots$, are defined by the following criterion:

IF $\alpha_k^{(2)} / \alpha_k^{(1)} \le \tau_k$ THEN
  $\alpha_k = \min\left\{\alpha_j^{(2)},\ j = \max\{1, k - M_\alpha\}, \ldots, k\right\}$;
  $\tau_{k+1} = \tau_k * 0.9$;
ELSE
  $\alpha_k = \alpha_k^{(1)}$; $\quad \tau_{k+1} = \tau_k * 1.1$;
ENDIF

where M_α is a prefixed non-negative integer and $\tau_1 \in (0,1)$. Concerning the choice of the scaling matrix D_k, a suited up-

dating rule generally depends on the special form of the objective function and then we discuss this topic later, when some applications of SGP will be presented.

3. TWO APPLICATIONS OF SGP

The numerical results reported in [4] show that SGP is a useful iterative regularization method for the deconvolution of images corrupted by Poisson noise (see also Section 4). These promising results motivate the application of SGP to other imaging problems, such as, for example, the deconvolution of images corrupted by Gaussian noise and the removal of Poisson noise.

3.1 Least-squares approach to nonnegative image deblurring

It is well-known that the maximum likelihood approach to image deblurring in the case of Gaussian noise leads to the following nonnegatively constrained least-squares problem:

$$\min_{x \geq 0} \quad J_y(x) = \tfrac{1}{2}\|Ax-y\|^2 \, .$$

Due to the ill-posedness of this convex problem, several iterative methods with the semiconvergence property have been proposed to obtain regularized solution by an early stopping of the iterations. Among these methods, the Projected Landweber (PL) method [7] and ISRA [6] are the most popular approaches. The PL iteration is given by

$$x^{(k+1)} = P_+(x^{(k)} - \alpha \nabla J_y(x^{(k)})) = P_+(x^{(k)} + \alpha A^T(y - Ax^{(k)})) \, ,$$

where α is a fixed steplength satisfying the conditions $0 < \alpha < \frac{2}{||A||^2}$, while ISRA can be described by

$$x^{(k+1)} \; = \; x^{(k)} \frac{A^T y}{A^T A x^{(k)}} \; = \; x^{(k)} - \frac{x^{(k)}}{A^T A x^{(k)}} \nabla J_y(x^{(k)}) \, ,$$

where the multiplication and the division of vectors are in the Hadamard sense, i.e. pixel by pixel. The interested reader can refer to the above mentioned references for more details on the two methods; here, we emphasize that SGP can be considered a generalization of both PL and ISRA. In fact, SGP is based on a gradient projection as PL, but it uses variable steplengths and scaled directions not exploited by PL. On the other hand, SGP can implement the same scaling of ISRA, but it benefits also from the variable steplengths not available in ISRA. Thus, in order to evaluate if the more general form can be useful in practice, it seems natural to update the SGP scaling matrix according to the scaling of ISRA:

$$D_k = \mathrm{diag}\left(\min\left[L, \, \max\left\{ \frac{1}{L}, \frac{x^{(k)}}{A^T A x^{(k)}} \right\} \right] \right) \, ,$$

where L is a large threshold ($L = 10^{10}$ in our simulations). For the sake of completeness and to better assess the importance of the scaling, we test also the behaviour of an SGP version without scaling matrix, that is with $D_k = I$ in each iteration; the corresponding method is denote by GP. The other main parameters of SGP and GP are set as follows:

- steplength parameter: $\alpha_{min} = 10^{-10}$, $\alpha_{max} = 10^5$, $\alpha_0 = 1.3$, $M_\alpha = 2$, $\tau_1 = 0.5$;
- line-search parameter: $\theta = 0.4$, $\beta = 10^{-4}$, $M = 1$.

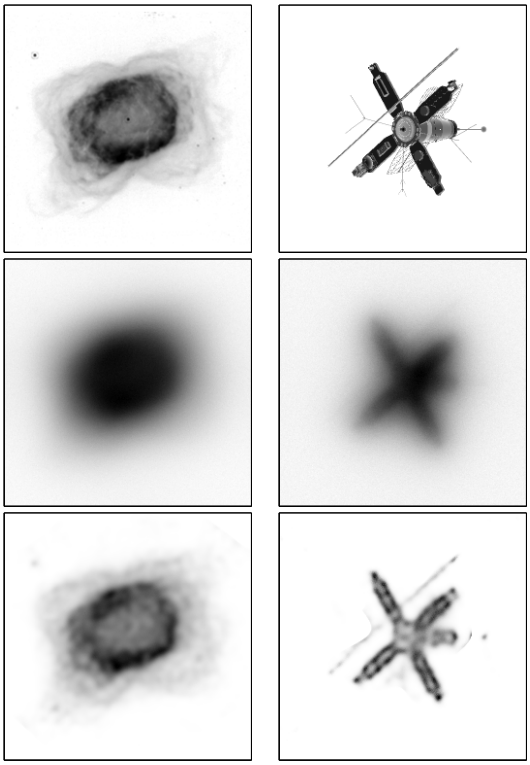

Figure 1: The original images (upper panels), the corresponding blurred and noisy images (middle panels) and the reconstructed images (lower panels).

We report the results obtained by comparing the schemes on two test problems generated from the 256×256 original images shown in the upper panels of Figure 1: an image of the nebula NGC5979 (left panel) and the frequently used satellite image (right panel). Blurred and noisy images (middle panels in Figure 1) are obtained by convolving these images with the point spread function available at http://www.mathcs.emory.edu/ nagy/RestoreTools/index.html and by perturbing the results with additive white Gaussian noise with zero expected value and variances $\sigma^2 = 1$. For each method, we evaluate the relative reconstruction error, defined as $\|x^{(k)} - x\|/\|x\|$, where x is the image to be reconstructed and $x^{(k)}$ is the reconstruction after k iterations; then we report the minimum relative reconstruction error (err), the number of iterations (it) and the computational time in seconds (sec) required to provide the minimum error. Table 1 summarizes the numerical results obtained by implementing the algorithms in standard C on an AMD Athlon 3500+ 2.2GHz processor; the same starting image is considered for all the schemes: $x_i^{(0)} = c, \; i = 1, \ldots, n, \;\; c = \sum_{j=1}^m y_j/n$. The SGP method is clearly preferable in terms of number of iterations and, even if it exhibits a cost per iteration larger than the other schemes, it generally provides the best overall computational time. The best SGP reconstructions are shown in the lower panels of Figure 1.

3.2 Edge-preserving removal of Poisson noise

We consider the optimization-based approach for removing Poisson noise recently proposed in [15], that consists in

Table 1: Image deblurring: behaviour of the algorithms.

Method	err	it	sec	err	it	sec
	Nebula NGC5979			Satellite		
PL	0.081	1134	7.90	0.278	10307	70.61
ISRA	0.075	1800	13.38	0.300	5248	38.42
SGP	0.075	90	1.88	0.302	314	6.28
GP	0.080	136	1.77	0.278	1768	21.51

the minimization of a functional obtained by penalizing the Kullback-Leibler divergence by means of an edge preserving functional. In order to introduce this optimization problem we use the following notation. Let $y \in \mathbb{R}^n$ be the detected noisy image; furthermore, if x_i denotes the value associated to the i-th pixel of the image x, we denote by x_{i^+} and x_{i^-} the values associated to the pixels that are just below and above the i-th pixel, respectively; similarly, x_{i_+} and x_{i_-} denote the values associated to the pixels that are just after and before the i-th pixel, respectively. Then, an estimate of the noise-free image can be obtained by solving the problem

$$\min_{x \geq \eta} \quad \sum_i \left(x_i - y_i - y_i \ln \frac{x_i}{y_i} \right) + \tau \left(\frac{1}{2} \sum_i \psi_\delta \left(\Delta_i^2 \right) \right), \qquad (2)$$

where $\Delta_i^2 = (x_{i_+} - x_i)^2 + (x_{i^+} - x_i)^2$ and $\psi_\delta(t) = 2\sqrt{t+\delta^2}$, δ being a small quantity tuning the discontinuities into the image [5], $\eta > 0$ is a small positive constant smaller than the background emission and $\tau > 0$ is a regularization parameter. See also [14] for a recent analysis of the total variation minimization. In the feasible region the objective function of the problem (2) is continuously differentiable, thus the SGP method can be applied. In particular, following the gradient splitting idea presented in [10], a diagonal scaling matrix for SGP can be defined as

$$(D_k)_{i,i} = \min \left(L, \ \max \left[\frac{1}{L}, \frac{x_i^{(k)}}{1 + \tau V_i(x^{(k)})} \right] \right), \qquad (3)$$

where $V_i(x^{(k)}) = \left\{ 2\psi'_\delta(\Delta_i^2) + \psi'_\delta(\Delta_{i_-}^2) + \psi'_\delta(\Delta_{i^-}^2) \right\} x_i^{(k)}$.
In this numerical experience we compare the SGP described in section 2, equipped with the scaling matrix (3), and two more traditional gradient algorithms. The first is a gradient projection (GP) method obtained by using $D_k = I$ in SGP. The second algorithm, that we denote by GP-BB, can be also derived by the SGP by setting $D_k = I$ and $\alpha_k = \alpha_k^{(1)}$; this scheme is very similar to the GPRS-BB method proposed in [8]. The numerical experiments have been carried out on two test problems generated by corrupting with Poisson noise the images in Figure 2. For each test problem we determine an approximate optimal value of the objective function by running the monotone SGP method until the relative difference between the function values at two successive iterates is less than 10^{-7}. Then we use this reference value as stopping criterion for the other algorithms. In Table 2 the results obtained on the LCR-phantom image are summarized in terms of ℓ_2 relative error, iterations number and computational time; in this case, Matlab implementations of the algorithms are run on an AMD Opteron Dual Core 2.4 GHz processor. We set $\eta = 10^{-5}$ and the regularization parameter τ equal to 0.25,

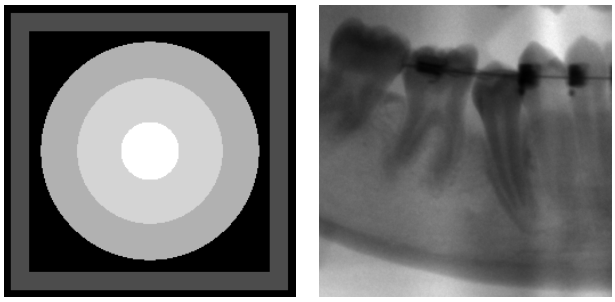

Figure 2: Original images for denoising problems: the LCR-phantom described in [15] (256×256) and a dental radiography (512×512).

Table 2: Image denoising: behaviour of the algorithms on the LCR-phantom.

Method	err	it	sec	err	it	sec
	$\delta = 10^{-8}$			$\delta = 10^{-1}$		
SGP	0.027	369	35.14	0.025	148	14.30
GP	0.032	2000*	159.13	0.025	280	23.23
GP-BB	0.032	2000*	156.35	0.025	735	70.62

a value derived from the literature for this test image. Looking at Table 2, we may observe that the SGP versions largely outperform the other algorithms, confirming the importance of both the scaling strategy and the alternating steplength selection. The SGP method has been applied also for the test problem obtained by perturbing with Poisson noise the dental radiography in the right panel of the Figure (2). In this case the relative difference in ℓ_2 norm between the noisy and the noise-free image is 0.179 and the parameters δ and τ has been estimated by an empirical procedure. The best results in terms of accuracy are obtained with $\delta = 10^{-1}$ and $\tau = 0.25$, for which the SGP method provides an approximation of the true object with an error equal to 0.028, in 113 iterations performed in 42 seconds.

4. SGP ON GRAPHICS PROCESSING UNITS

In this section we discuss an implementation of SGP for Graphics Processing Units (GPU), that are non-expensive parallel processing devices available on many up-to-date personal computers. For this experience we apply SGP as iterative regularization algorithm for the deblurring of images corrupted by Poisson noise, that is, for an approximate solution of

$$\min_{x \geq 0} \ J_y(x) = \sum_{i=1}^m \left(\sum_{j=1}^n A_{ij} x_j + b - y_i - y_i \ln \frac{\sum_{j=1}^n A_{ij} x_j + b}{y_i} \right),$$

where $b > 0$ denotes a constant background radiation. An effective version of SGP for such kind of application has been recently proposed in [4] by deriving the updating rule for the matrix D_k from the scaling used by the well-known EM method [12]. Taking into account that, under standard assumption on the blurring matrix A, the EM method for the above minimization problem can be written as

$$x^{(k+1)} = x^{(k)} \frac{A^T y}{A x^{(k)} + b} = x^{(k)} - x^{(k)} \nabla J_y(x^{(k)}),$$

the following choice for the matrix D_k is suggested in [4]:

$$D_k = \mathrm{diag}\left(\min\left[L,\ \max\left\{\frac{1}{L}, x^k\right\}\right]\right) .$$

Thus, SGP can be viewed as a generalization of the EM method able to exploit variable steplengths for improving the convergence rate. The fast convergence allowed SGP to largely improve the EM computational time on the standard serial architecture tested in [4]. Here, we show that the same holds when these algorithms are implemented on modern graphics processors. In Table 3 we report the numerical results obtained by running EM and SGP on a NVIDIA GTX 280 graphics card, equipped with 30 streaming multiprocessors (240 total cores) running at 1296 MHz and with 1GB of global memory. The parallel implementation is developed within the programming environment CUDA (Compute Unified Device Architecture), provided by NVIDIA (see http://www.nvidia.com/cuda). In particular, the kernel libraries of CUDA called CUFFT and CUBLAS are exploited for performing the 2-D FFTs and the linear algebra operations. For evaluating speedups in comparison with standard CPU implementations, we report also the results obtained by means of a C implementation (within the Microsoft Visual Studio 2005 environment) on a personal computer equipped with an AMD Athlon X2 Dual-Core at 3.11GHz and 3GB of RAM. The test problems we consider are generated as in [4, 13]: an image of the nebula NGC5979 sized 256×256 is convolved with an ideal PSF, then a constant background term is added and the resulting image is perturbed with Poisson noise. Test problems of larger size are obtained by expanding the original images and the PSF by means of a zero-padding technique on their FFTs. The same parameters given in Section 3.1 are used for SGP, except the line-search parameter M, for which the setting $M = 10$ is preferred. The results in Table 3 show that the GPU implementations allow us to save time over the CPU implementation for more than one order of magnitude. Furthermore, SGP largely outperforms EM also on graphics processors, even if the additional operations required in each SGP iteration (in particular the scalar products involved in the steplength selection) make this algorithm less suited than EM for a parallel implementation on GPU. We refer to [13] for a deeper analysis of the SGP implementation on GPU and to [11] for examples of other image reconstruction algorithms implemented on graphics processors.

Table 3: Image deblurring: SGP and EM on GPU.

	CPU		GPU		
n	err	sec	err	sec	speedup
SGP (it = 29)					
256^2	0.070	0.72	0.071	0.05	14.7
512^2	0.065	2.69	0.065	0.16	16.8
1024^2	0.064	10.66	0.064	0.58	18.4
2048^2	0.064	49.81	0.063	2.69	18.5
EM (it = 500)					
256^2	0.070	4.41	0.071	0.19	23.2
512^2	0.064	19.91	0.064	0.89	22.4
1024^2	0.063	97.92	0.063	3.63	27.0
2048^2	0.063	523.03	0.063	23.05	22.7

REFERENCES

[1] J. Bardsley and J. Nagy, "Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging," *SIAM J. Matr. Anal. Appl.*, vol. 27, pp. 1184–1198, 2006.

[2] J. Barzilai and J. M. Borwein, "Two point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, pp. 141–148, 1988.

[3] E. Birgin, J. Martinez, and M. Raydan, "Inexact spectral projected gradient methods on convex sets," *IMA J. Numer. Anal.*, vol. 23, pp. 539–559, 2003.

[4] S. Bonettini, R. Zanella, and L. Zanni, "A scaled gradient projection method for constrained image deblurring," *Inverse Problems*, vol. 25, 015002, 2009.

[5] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and A. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, pp. 298-311, 1997.

[6] M. E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorithm suitable for volume ECT," *IEEE Trans. Med. Imaging*, vol. 5, pp. 61-66, 1986.

[7] B. Eicke, "Iteration methods for convexly constrained ill-posed problems in Hilbert spaces," *Numer. Funct. Anal. Optim.*, vol. 13, pp. 413-29, 1992.

[8] M. A. T. Figueriredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics in Sig. Proc.*, vol. 1, pp. 586-597, 2007.

[9] G. Frassoldati, L. Zanni, and G. Zanghirati, "New adaptive stepsize selections in gradient methods," *J. Indust. Manag. Optim.*, vol. 4, pp. 299-312, 2008.

[10] H. Lanteri, M. Roche, and C. Aime, "Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms," *Inverse Problems*, vol. 18, pp. 1397–1419, 2002.

[11] S. Lee and S. J. Wright "Implementing algorithms for signal and image reconstruction on graphical processing units.", submitted 2008. Available at: http://www.optimization-online.org/DB_HTML/2008/11/2131.html

[12] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. 1, pp. 113–122, 1982.

[13] V. Ruggiero, T. Serafini, R. Zanella, and L. Zanni, "Iterative regularization algorithms for constrained image deblurring on graphics processors,", *J. of Global Optimization*, to appear, 2009.

[14] P. Weiss, L. Blanc-Féraud and G. Aubert, "Efficient schemes for total variation minimization under constraints in image processing", *SIAM J. Sci. Comput.*, vol. 31, pp. 2047-2080, 2009.

[15] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, "Efficient gradient projection methods for edge–preserving removal of Poisson noise," *Inverse Problems*, vol. 25, 045010, 2009.