PRIMAL-DUAL FIRST ORDER METHODS FOR TOTAL VARIATION IMAGE RESTORATION IN PRESENCE OF POISSON NOISE

Silvia Bonettini, Alessandro Benfenati, Valeria Ruggiero

Dipartimento di Matematica e di Informatica, Università di Ferrara
Polo Scientifico Tecnologico, Blocco B, Via Saragat 1, I-44122 Ferrara, Italy

ABSTRACT

Image restoration often requires the minimization of a convex, possibly nonsmooth functional, given by the sum of a data fidelity measure plus a regularization term. In order to face the lack of smoothness, alternative formulations of the minimization problem could be exploited via the duality principle. Indeed, the primal-dual and the dual formulation have been well explored in the literature when the data suffer from Gaussian noise and, thus, the data fidelity term is quadratic. Unfortunately, the most part of the approaches proposed for the Gaussian are difficult to apply to general data discrepancy terms, such as the Kullback-Leibler divergence. In this work we propose primal-dual methods which apply to the minimization of the sum of general convex functions and whose iteration is easy to compute, regardless of the form of the objective function, since it essentially consists in a subgradient projection step. We provide the convergence analysis and we suggest some strategies to improve the convergence speed by means of a careful selection of the steplength parameters. A numerical experience on Total Variation based denoising and deblurring problems from Poisson data shows the behavior of the proposed method with respect to other state-of-the-art algorithms.

***Index Terms*—** Primal–Dual method, ϵ–subgradient projection method, variable steplengths, Total Variation, Kullback–Leibler divergence.

1. INTRODUCTION

In the Bayesian framework, image reconstruction problems can be formulated as a constrained convex minimization problems of the form

$$\min_{\mathbf{x}\in X} f(\mathbf{x}) \tag{1}$$

where $X \subseteq \mathbb{R}^n$ is the constraints set and the objective function can be written as

$$f(\mathbf{x}) = f_0(\mathbf{x}) + f_1(A\mathbf{x})$$

The function $f_0(\mathbf{x})$ measures the data discrepancy and should be chosen according to the noise statistics: in particular, when the data suffer from Poisson noise, the Maximum Likelihood principle leads to the Kullback–Leibler divergence

$$f_0(\mathbf{x}) = \sum_{i=1}^{n} g_i \log \frac{g_i}{(H\mathbf{x})_i + b} + (H\mathbf{x})_i + b - g_i \tag{2}$$

where $\mathbf{g} \in \mathbb{R}^n$ is the observed image, $H \in \mathbb{R}^{n\times n}$ is the imaging matrix representing the image acquisition system and b is a nonnegative constant background term. On the other side, $f_1(A\mathbf{x})$, where $A \in \mathbb{R}^{m\times n}$, is a regularization term which enforces suitable properties on the solution of (1). Typically, to preserve the edges in the solutions of (1), $f_1(A\mathbf{x})$ represents the discrete, nonsmooth, Total Variation (TV) functional

$$f_1(A\mathbf{x}) = \beta \sum_{i=1}^{n} \|A_i\mathbf{x}\|, \quad A_i \in \mathbb{R}^{d\times n} \tag{3}$$

where β is a positive regularization parameter, $\|\cdot\|$ denotes the Euclidean norm and $A = \begin{pmatrix} A_1^T & A_2^T & \cdots & A_n^T \end{pmatrix}^T$ represents the first order finite difference operator. In order to face the lack of smoothness in the objective function of (1), the primal–dual formulation can be introduced via the duality principle, yielding the following saddle point problem

$$\min_{\mathbf{x}\in X} \max_{\mathbf{y}} f_0(\mathbf{x}) + \langle A\mathbf{x}, \mathbf{y}\rangle - f_1^*(\mathbf{y}) \tag{4}$$

where $\langle \cdot, \cdot \rangle$ represents the inner product of two m vectors and f_1^* is the convex conjugate of f_1, i.e. $f_1^*(\mathbf{y}) = \max_{\mathbf{x}} \langle \mathbf{x}, \mathbf{y}\rangle - f_1(\mathbf{x})$. A key property of $f_1^*(\mathbf{y})$, which is crucial for our analysis, is the boundedness of its domain.

By defining the dual function $d(\mathbf{y}) = \min_{x\in X} f_0(\mathbf{x}) + \langle A\mathbf{x}, \mathbf{y}\rangle - f_1^*(\mathbf{y})$, the dual formulation of (1) can be derived: $\max_{\mathbf{y}} d(\mathbf{y})$. Primal-dual and dual formulations with the related methods have been deeply analyzed in the recent literature for the Gaussian noise case (see for example [1, 2, 3, 4, 5, 6]). However, the most part of such methods can not be directly applied when the data discrepancy is the Kullback–Leibler divergence. Indeed, when $f_0(\mathbf{x})$ is defined as in (2), the difficulty is that there is no explicit expression

This research is supported by the project *Learning meets time: A new computational approach for learning in dynamic systems*, contract RBFR12M3A, funded by the Italian Ministry of Education, University and Research *Futuro in Ricerca 2012* program. The National Group of Computing Science (GNCS-INDAM) is also acknowledged.

for the dual function and, except in case of denoising problems when H reduces to the identity matrix, there is no closed formula to compute the resolvent operator of f_0

$$(I + \theta\partial f_0)^{-1}(\mathbf{x}) = \arg\min_{\mathbf{z}} f_0(\mathbf{z}) + \frac{1}{2\theta}\|\mathbf{z} - \mathbf{x}\|^2$$

for given $\mathbf{x} \in X$, $\theta \in \mathbb{R}$, $\theta > 0$.

For the reasons above, in this paper we consider the saddle point problem (4), focusing on a primal–dual algorithm whose iteration requires only to compute an approximate subgradient of $f_0(\mathbf{x})$. This method has been proposed in our previous work [7] as a generalization of the hybrid gradient method proposed in [4] for the Gaussian case, providing the related convergence analysis. In this paper we state a stronger convergence result under very similar assumptions, providing also a convergence rate estimate.

1.1. Notation and basic definitions

Here and in the following $\|\cdot\|$ denotes the Euclidean norm of a vector and $\langle\cdot,\cdot\rangle$ the corresponding inner product. Moreover, $P_X(\mathbf{z})$ indicates the orthogonal projection of $\mathbf{z} \in \mathbb{R}^n$ onto the set X, i.e. $P_X(\mathbf{z}) = \arg\min_{\mathbf{x}\in X} \|\mathbf{x} - \mathbf{z}\|$ and the diameter of a set X is denoted by $\text{diam}(X) = \max_{\mathbf{x},\mathbf{z}\in X} \|\mathbf{x} - \mathbf{z}\|$.

The domain of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is the set

$$\text{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \neq \infty\}$$

We recall also the definition of the ϵ-subdifferential of a convex function f at x, which is the following set

$$\partial_\epsilon f(\mathbf{x}) = \{\mathbf{u} \in \mathbb{R}^n : f(\mathbf{z}) - f(\mathbf{x}) \geq \langle \mathbf{u}, \mathbf{z} - \mathbf{x}\rangle - \epsilon \; \forall \mathbf{z} \in \mathbb{R}^n\} \quad (5)$$

for a given $\epsilon \geq 0$. An ϵ-subgradient of f at x is any element of the set $\partial_\epsilon f(\mathbf{x})$. Due to obvious space limitation reasons we address the reader to [8, 9] for the basic properties of the ϵ-subgradient.

The ϵ-subgradient projection method [10, 11, 12] for the solution of problem (1) consists in the following iteration

$$\mathbf{x}^{(k+1)} = P_X(\mathbf{x}^{(k)} - \theta_k \mathbf{u}^{(k)}) \quad (6)$$

where $\mathbf{u}^{(k)} \in \partial_{\epsilon_k} f(\mathbf{x}^{(k)})$, for suitably chosen $\epsilon_k \geq 0$ and $\theta_k > 0$.

2. ALGORITHM DESCRIPTION

The primal–dual hybrid gradient (PDHG) algorithm [7] consists in the following iteration

$$\mathbf{y}^{(k+1)} = (I + \tau_k \partial f_1^*)^{-1}(\mathbf{y}^{(k)} + \tau_k A\mathbf{x}^{(k)}) \quad (7)$$
$$\mathbf{x}^{(k+1)} = P_X(\mathbf{x}^{(k)} - \theta_k(\mathbf{g}^{(k)} + A^T\mathbf{y}^{(k+1)})) \quad (8)$$

where $\mathbf{x}^{(0)} \in X$, $\mathbf{y}^{(0)} \in \text{dom}(f_1^*)$, $\mathbf{g}^{(k)} \in \partial_{\delta_k} f_0(\mathbf{x}^{(k)})$, $\delta_k \geq 0$ and $\{\theta_k\}$, $\{\tau_k\}$ are given primal and dual positive steplength sequences. The PDHG method has been proposed first in [4] to solve unconstrained Total Variation based image restoration problems in presence of Gaussian noise and its very good practical performances have been observed by several authors [3, 2]. As observed in [4], the main strength of PDHG resides in exploiting variable steplengths θ_k and τ_k: in particular, using a large primal stepsize in the initial iterates, decreasing its value as the iterations proceed, while adopting increasing dual stepsizes, leads to obtain a fast initial and asymptotic approaching to an optimal solution.

It is worth stressing that in case of Total Variation regularization, the practical implementation of the dual updating formula (7) reduces to an orthogonal projection of the vector $\mathbf{y}^{(k)} + \tau_k A\mathbf{x}^{(k)}$ onto a cartesian product of n Euclidean balls, which is very easy to compute. On the other side, when $f_0(\mathbf{x})$ is differentiable, one may choose $\mathbf{g}^{(k)} = \nabla f_0(\mathbf{x}^{(k)})$ in (8), while the primal domain X is typically defined as the nonnegative orthant of $\mathbb{R}^n$, $X = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq 0\}$, so that the projection is straightforward. More general operators are considered for example in [13].

A general convergence analysis of PDHG has been developed in [7], where, under suitable assumptions, it is proved that the distance between the iterates and the set of the solutions of (1) converges to zero. The key point of this analysis consists in the interpretation of the scheme (7)–(8) as a special case of an ϵ-subgradient projection method. The basic result for this interpretation is the following lemma.

Lemma 1 *[7] Let $\mathbf{y}^{(k+1)}$ defined as in (7). Then, $A^T\mathbf{y}^{(k+1)} \in \partial_{\sigma_k}(f_1 \circ A)(\mathbf{x}^{(k)})$, where $\sigma_k = f_1(A\mathbf{x}^{(k)}) + f_1^*(\mathbf{y}^{(k+1)}) - \langle A\mathbf{x}^{(k)}, \mathbf{y}^{(k+1)}\rangle$. If, in addition, there exists a positive number D such that diam(domf_1^*) $< D$, then $\sigma_k \leq (2\tau_k)^{-1}D^2$.*

As a consequence of the previous result, we obtain that

$$\mathbf{u}^{(k)} = \mathbf{g}^{(k)} + A^T\mathbf{y}^{(k+1)} \in \partial_{\epsilon_k} f(\mathbf{x}^{(k)}) \quad (9)$$

where $\epsilon_k = \delta_k + \sigma_k$. Following this idea, in the next section we prove the convergence of the whole sequence generated by PDHG to a solution of (1), providing also a convergence rate estimate.

3. CONVERGENCE ANALYSIS

For the ϵ-subgradient projection method (6) the following convergence result holds .

Theorem 1 *[10, Theorem 8] Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by iteration (6) applied to problem (1). Assume that the set of the solutions of (1) is nonempty and bounded and that $\lim_{k\to\infty} \epsilon_k = 0$. Moreover, assume that there exists $\rho > 0$ such that $\|\mathbf{u}^{(k)}\| \leq \rho$ for all k and the steplength sequence satisfies*

$$\sum_{k=0}^{\infty} \theta_k = \infty, \quad \sum_{k=0}^{\infty} \theta_k^2 < \infty, \quad \sum_{k=0}^{\infty} \epsilon_k\theta_k < \infty \quad (10)$$

Then, the sequence $\{\mathbf{x}^{(k)}\}$ converges to a solution of (1).

Based on the previous result and on (9), the convergence of the primal–dual scheme (7)–(8) can be stated as follows.

Corollary 1 *Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by iteration* (7)-(8) *applied to problem* (4). *Assume that there exists $\rho > 0$ such that $\|\mathbf{g}^{(k)}\| \leq \rho$ for all k and the steplength sequence satisfies*

$$\theta_k = \mathcal{O}\left(\frac{1}{k^p}\right), \quad \tau_k = \mathcal{O}(k^p) \quad \frac{1}{2} < p \leq 1. \tag{11}$$

Moreover, assume that the set of the solutions of (1) *is nonempty and bounded and that δ_k converges to zero at least as $\frac{1}{\tau_k}$. Then, the sequence $\{\mathbf{x}^{(k)}\}$ converges to a solution of* (1).

Under the same assumptions of Theorem 1, we can give the following convergence rate estimate, whose proof is based on the results in [11].

Theorem 2 *Let $\{\mathbf{x}^{(k)}\}$ be the sequence generated by iteration* (6). *Assume that the hypotheses of Theorem 1 are satisfied. Then, there exists a subsequence $\{\mathbf{x}^{(\ell_k)}\}$ of $\{\mathbf{x}^{(k)}\}$ such that*

$$f(\mathbf{x}^{(\ell_k)}) - f(\mathbf{x}^*) \leq \left(\sum_{j=0}^{\ell_k} \theta_j\right)^{-1} \tag{12}$$

where $\mathbf{x}^*$ is a solution of (1).

Proof. For all k let us define $\mathbf{z}^{(k)} = \mathbf{x}^{(k)} - \theta_k \mathbf{u}^{(k)}$. By the properties of the projection operator we have that

$$\begin{aligned} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| &= \|P_X(\mathbf{z}^{(k)}) - P_X(\mathbf{x}^{(k)})\| \\ &\leq \|\mathbf{z}^{(k)} - \mathbf{x}^{(k)}\| \qquad (13) \\ &= \theta_k \|\mathbf{u}^{(k)}\| \leq \theta_k \rho \qquad (14) \end{aligned}$$

Let $\mathbf{x}$ be any solution of (1): then, we have

$$\begin{aligned} &\theta_k^2 \rho^2 + \|\mathbf{x}^{(k)} - \mathbf{x}\|^2 - \|\mathbf{x}^{(k+1)} - \mathbf{x}\|^2 \\ &\geq \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 + \|\mathbf{x}^{(k)} - \mathbf{x}\|^2 - \|\mathbf{x}^{(k+1)} - \mathbf{x}\|^2 \\ &= 2\langle \mathbf{x}^{(k)} - \mathbf{x}, \mathbf{x}^{(k)} - \mathbf{x}^{(k+1)} \rangle \\ &= 2\langle \mathbf{x}^{(k)} - \mathbf{x}, \mathbf{x}^{(k)} - \mathbf{z}^{(k)} \rangle + 2\langle \mathbf{x}^{(k)} - \mathbf{x}, \mathbf{z}^{(k)} - \mathbf{x}^{(k+1)} \rangle \\ &= 2\theta_k \langle \mathbf{u}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x} \rangle + 2\langle \mathbf{x}^{(k)} - \mathbf{z}^{(k)}, \mathbf{z}^{(k)} - \mathbf{x}^{(k+1)} \rangle + \\ &\quad +2\langle \mathbf{z}^{(k)} - \mathbf{x}, \mathbf{z}^{(k)} - \mathbf{x}^{(k+1)} \rangle \\ &= 2\theta_k \langle \mathbf{u}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x} \rangle + 2\langle \mathbf{x}^{(k)} - \mathbf{z}^{(k)}, \mathbf{z}^{(k)} - \mathbf{x}^{(k+1)} \rangle + \\ &\quad +2\langle \mathbf{z}^{(k)} - \mathbf{x}, \mathbf{z}^{(k)} - P_X(\mathbf{z}^{(k)}) \rangle \\ &\geq 2\theta_k \langle \mathbf{u}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x} \rangle + 2\langle \mathbf{x}^{(k)} - \mathbf{z}^{(k)}, \mathbf{z}^{(k)} - \mathbf{x}^{(k+1)} \rangle \\ &= 2\theta_k \langle \mathbf{u}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x} \rangle + 2\langle \mathbf{x}^{(k)} - \mathbf{z}^{(k)}, \mathbf{z}^{(k)} - \mathbf{x}^{(k)} \rangle \\ &\quad +2\langle \mathbf{x}^{(k)} - \mathbf{z}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x}^{(k+1)} \rangle \\ &\geq 2\theta_k \langle \mathbf{u}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x} \rangle - 2\|\mathbf{x}^{(k)} - \mathbf{z}^{(k)}\|^2 + \\ &\quad -2\|\mathbf{x}^{(k)} - \mathbf{z}^{(k)}\| \|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\| \\ &\geq 2\theta_k \langle \mathbf{u}^{(k)}, \mathbf{x}^{(k)} - \mathbf{x} \rangle - 4\theta_k^2 \|\mathbf{u}^{(k)}\|^2 \\ &\geq 2\theta_k (f(\mathbf{x}^{(k)}) - f(\mathbf{x}) - \epsilon_k) - 4\theta_k^2 \rho^2 \end{aligned}$$

By defining $\gamma_k = f(\mathbf{x}^{(k)}) - f(\mathbf{x})$, this results in

$$2\theta_k \gamma_k \leq \|\mathbf{x}^{(k)} - \mathbf{x}\|^2 - \|\mathbf{x}^{(k+1)} - \mathbf{x}\|^2 + 2\theta_k \epsilon_k + 5\theta_k^2 \rho^2 \tag{15}$$

By summing inequalities (15) for $k = 0, 1, \cdots, N$ we obtain

$$2\sum_{k=0}^{N} \theta_k \gamma_k \leq \|\mathbf{x}^0 - \mathbf{x}\|^2 + +2\sum_{k=0}^{N} \theta_k \epsilon_k + 5\rho^2 \sum_{k=0}^{N} \theta_k^2$$

which implies that $\sum_{k=0}^{\infty} \theta_k \gamma_k < \infty$. Then, proceeding as in the proof of Theorem 2 in [11], the estimate (12) follows. □

The previous result gives a quite pessimistic convergence rate estimate of the objective function value: indeed, when the primal steplength is chosen as in (11) with $p = 1$, we can only say that there exists a subsequence of $\{f(\mathbf{x}^{(\ell_k)})\}$ of $\{f(\mathbf{x}^{(k)})\}$ such that

$$f(\mathbf{x}^{(\ell_k)}) - f(\mathbf{x}^*) \leq \frac{1}{\log(\ell_k)}$$

However, as illustrated in the next section, the practical performances are significantly better than this theoretical estimate.

	method	K	time(s.)	e^K
		$f^K < 100$		
LCR	CP	110	4.36	0.0027468
	PDHG	188	4.97	0.0005846
	PID	119	7.48	0.0013933
		$f^K < 1$		
	method	K	time(s.)	e^K
	CP	1652	62.83	1.2329e-4
	PDHG	943	26.06	8.9046e-6
	PID	398	22.68	8.3888e-5
		$f^K < 100$		
	method	K	time(s.)	e^K
micro	PID, $\gamma = \frac{50}{\beta}$	36	1.37	0.042813
	PID, $\gamma = \frac{5}{\beta}$	101	4.06	0.024150
	PDHG	865	7.19	0.021812
		$f^K < 1$		
	method	K	time(s.)	e^K
	PID, $\gamma = \frac{50}{\beta}$	3509	134.03	0.011424
	PID, $\gamma = \frac{5}{\beta}$	382	14.82	0.011232
	PDHG	2090	17.17	0.013842

Table 1. Numerical results.

4. NUMERICAL EXPERIENCE

This section is devoted to numerically evaluate the behavior of PDHG method for TV restoration of images corrupted by

Poisson noise. The numerical experiments have been performed in Matlab environment, on a server with a dual Intel Xeon QuadCore E5620 processor at 2,40 GHz, 12 Mb cache and 18 Gb of RAM.
We consider a denoising and a deblurring problem, denoted by *LCR* and *micro* respectively. In both cases the Poisson noise has been simulated by the Matlab `imnoise` function. The features of the two test problems are the following:

- *LCR*: the original image is an array 256×256 [14], consisting in concentric circles of intensities 70, 135 and 200, enclosed by a square frame of intensity 10, all on a background of intensity 5; β is equal to 0.25;
- *micro*: the original image is the confocal microscopy phantom of size 128×128 [15]; its values are in the range $[0, 70]$ and the total flux is $2.9461\ 10^5$; the simulated data are obtained by a Gaussian PSF and the background term b in (2) is set to zero; β is equal to 0.09.

For both test-problems we compute the solution x^* of (1)-(2)-(3) by running 100000 iterations of the PIDSplit method [16]. Then, we evaluate the progress toward the ideal solution at each iteration in terms of the l_2 relative error $e^k = \frac{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|}{\|\mathbf{x}^*\|}$ and of the distance $f^k = f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*)$ from the minimum value.
For the denoising test-problem *LCR*, we compare PDHG with two methods: the first one is PIDSplit algorithm [16, 17], based on a very efficient alternating direction multiplier method and depending on a positive parameter γ; the second one is the Algorithm 1 in [2] (CP), a general primal–dual scheme which depends on two parameters, σ and τ, satisfying $\sigma\tau L^2 < 1$, with $\beta\|A\| \leq L$. Since applying CP to the deblurring case without approximating the resolvent operators with an inner loop requires a reformulation of the primal problem (see [18, 19, 20]), we apply it only to the denoising problem. All methods have been initialized with $\mathbf{x}^{(0)} = \max(\mathbf{g}, \eta)$, where the maximum is intended componentwise and $\eta_i = 0$ for $g_i = 0$, $\eta_i = \min_{g_i > 0}(\mathbf{g})$ otherwise. The initial guess for the dual variables has been set equal to zero. By a *trial* and *error* procedure, we choose for the three methods optimized parameters: in particular for PDHG the optimized sequences for the steplengths are $\tau_k = 0.4 + 0.01k$ and $\theta_k = \frac{1}{0.15 + 0.0015k}$ while for PIDSplit $\gamma = \frac{1}{\beta}$ and for CP $\tau = 5$. In Figure 1 we show in log-scale the relative error e^k and the distance from the minimum value f^k with respect to the computational time for 3000 iterations of PDHG , PIDSplit and CP.
For the deblurring test-problem *micro*, we compare the results obtained by PDHG and PIDSplit methods. The initial setting is the same of the denoising, except for the primal variable equal to $\mathbf{x}^{(0)} = \max(\mathbf{g}, 0)$. The *optimal* sequences for PDHG are $\tau_k = 0.9 + 0.01k$, $\theta_k = \frac{1}{0.33 + 2\ 10^{-4}k}$ while for PIDSplit we use two values for γ, i. e. $\frac{5}{\beta}$ and $\frac{50}{\beta}$. Figure 2 show in log–scale the relative error e^k and the distance f^k

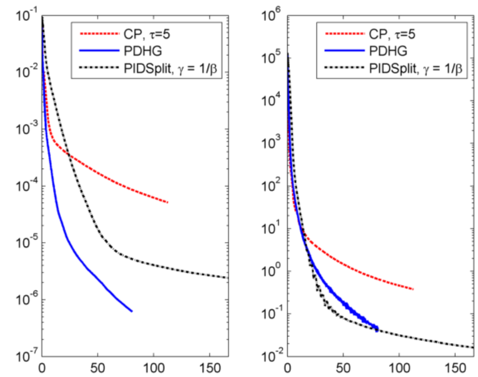

Fig. 1. Test-problem *LCR*: behavior of e^k (left) and f^k (right) with respect the computational time in seconds.

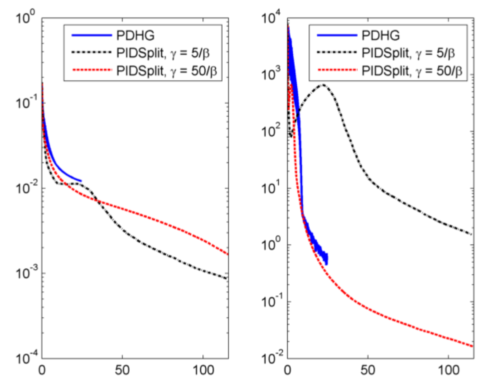

Fig. 2. Test-problem *micro*: behavior of e^k (left) and f^k (right) with respect the computational time in seconds.

from the minimum value with respect to the computational time for 3000 iterations of PDHG and PIDSplit. In Table 1 we report the number of iterations K and the computational time needed to satisfy $f^K < \mu$, for $\mu = 100$ and $\mu = 1$, and the corresponding relative error e^K for both test-problems.
We observe that the behavior of PDHG is comparable to other methods; the choice of the steplength sequences is crucial for its effectiveness, as well as the choice of γ for PIDSplit and of τ for CP. Furthermore the comparison between the methods shows that a fast convergence for the relative error e^k often does not match with an analogous behavior for f^k, because of the different trajectories to go from $\mathbf{x}^{(0)}$ to $\mathbf{x}^*$.

5. REFERENCES

[1] A. Chambolle, “An algorithm for Total Variation minimization and applications,” *J. Math. Imag. Vis.*, vol. 20, pp. 89–97, 2004.

[2] A. Chambolle and T. Pock, “A first–order primal–dual algorithm for convex problems with applications to imaging,” *J. Math. Imaging Vis.*, vol. 40, pp. 120–145, 2011.

[3] E. Esser, X. Zhang, and T. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1015–1046, 2010.

[4] M. Zhu and T. F. Chan, "An efficient primal–dual hybrid gradient algorithm for Total Variation image restoration," CAM Report 08-34, UCLA, 2008.

[5] M. Zhu, S. J. Wright, and T. F. Chan, "Duality-based algorithms for total-variation-regularized image restoration," *Computational Optimization and Applications*, vol. 47, pp. 377–400, 2008.

[6] J.-F. Aujol, "Some first-order algorithms for total variation based image restoration," *J. Math. Imaging Vis.*, vol. 34, no. 3, pp. 307–327, 2009.

[7] S. Bonettini and V. Ruggiero, "On the convergence of primal-dual hybrid gradient algorithms for total variation image restoration," *J. Math. Imaging Vis.*, vol. 44, pp. 236–253, 2012.

[8] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, SIAM, 1999.

[9] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[10] T. Larsson, M. Patriksson, and A.-B. Strömberg, "On the convergence of conditional ϵ-subgradient methods for convex programs and convex–concave saddle-point problems," *European Journal of Operational Research*, vol. 151, pp. 461–473, 2003.

[11] Ya.I. Alber, A.N. Iusem, and M.V. Solodov, "On the projected subgradient method for nonsmooth convex optimization in a Hilbert space," *Math. Prog.*, vol. 81, pp. 23–35, 1998.

[12] S. M. Robinson, "Linear convergence of epsilon-subgradient descent methods for a class of convex functions," *Math. Program., Ser. A*, vol. 86, 1999.

[13] P.L. Combettes and B.C. Vũ, "Variable metric forward-backward splitting with applications to monotone inclusions in duality," *Optimization*, to appear.

[14] T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by Poisson noise," *J. Math. Imaging Vis.*, vol. 27, pp. 257–263, 2007.

[15] R. M. Willett and R. D. Nowak, "Platelets: A multiscale approach for recovering edges and surfaces in photon limited medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 332–350, 2003.

[16] S Setzer, G Steidl, and T Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Vis. Commun. Image R.*, vol. 21, pp. 193–199, 2010.

[17] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.

[18] S. Anthoine, J.-F. Aujol, Y. Boursier, and C. Mélot, "Some proximal methods for CBCT and PET tomography," *Inverse Problems and Imaging*, vol. 6, no. 4, pp. 565–598, 2012.

[19] F.-X. Dupé, M. J. Fadili, and J.-L. Starck, "Inverse problems with Poisson noise: primal and primal-dual splitting," in *18th IEEE International Conference on Image Processing (ICIP), 2011*, Brussels, 2011, pp. 1901–1904, IEEE.

[20] F.-X. Dupé, M. J. Fadili, and J.-L. Starck, "Linear inverse problems with various noise models and mixed regularizations," in *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools*, Brussels, 2011, pp. 621–626, ICST.