INERTIAL VARIABLE METRIC TECHNIQUES FOR THE INEXACT FORWARD–BACKWARD ALGORITHM[*]

S. BONETTINI[†], S. REBEGOLDI[‡], AND V. RUGGIERO[‡]

Abstract. One of the most popular approaches for the minimization of a convex functional given by the sum of a differentiable term and a nondifferentiable one is the forward-backward method with extrapolation. The main reason making this method very appealing for a wide range of applications is that it achieves a $\mathcal{O}(1/k^2)$ convergence rate in the objective function values, which is optimal for a first order method. Recent contributions on this topic are related to the convergence of the iterates to a minimizer and the possibility of adopting a variable metric in the proximal step. Moreover, it has been also proved that the objective function convergence rate is actually $o(1/k^2)$. However, these results are obtained under the assumption that the minimization subproblem involved in the backward step is computed exactly, which is clearly not realistic in a variety of relevant applications. In this paper, we analyze the convergence properties when both variable metric and inexact computation of the backward step are allowed. To do this, we adopt a suitable inexactness criterion and we devise implementable conditions on both the accuracy of the inexact backward step computation and the variable metric selection, so that the $o(1/k^2)$ rate and the convergence of the iterates are preserved. The effectiveness of the proposed approach is also validated with a numerical experience showing the effects of the combination of inexactness with variable metric techniques.

Key words. Convex optimization, inertial forward–backward algorithms, inexact proximal operator, variable metric, image restoration.

AMS subject classifications. 65K05, 90C25, 90C30.

1. Introduction.

Optimization problems of the form

$$\min_{x\in\mathcal{H}} F(x) \equiv f(x) + g(x)\,, \tag{1.1}$$

where f and g are real-valued convex functions defined on a Hilbert space $\mathcal{H}$, are relevant in a variety of frameworks, such as signal and image restoration, machine learning, statistical inference. Typically, in such kind of applications, one of the two terms, say f, represents the data misfit and consists in a differentiable function, while the other one is included in the model to regularize the solution, i.e., to impose some desired properties, and it is often nonsmooth.

The class of forward–backward (FB) methods [5, 14] specifically addresses problem (1.1) by iteratively applying the following step

$$\operatorname{prox}_{\alpha_k g}(x^{(k)} - \alpha_k\nabla f(x^{(k)})) \equiv \arg\min_{x\in\mathcal{H}} g(x) + \frac{1}{2\alpha_k}\|x - x^{(k)} + \alpha_k\nabla f(x^{(k)})\|^2, \tag{1.2}$$

which combines a gradient (forward, explicit) step on the smooth part f with a proximal (backward, implicit) one related to the nonsmooth term g. FB methods include as special cases several popular algorithms such as the gradient projection algorithm [9], the Iterative Soft Tresholding Algorithm (ISTA) [5], and several variants of them [2, 32, 25, 24, 28].

In the last ten years, a great attention has been dedicated to the introduction of an *inertial (or extrapolation) step* in FB schemes, devising the so-called *inertial (or accelerated) first order methods.* One of the most well–known inertial approaches is based on the following variant of step (1.2):

$$\operatorname{prox}_{\alpha_k g}(y^{(k)} - \alpha_k\nabla f(y^{(k)})), \quad \text{where} \quad y^{(k)} = x^{(k)} + \beta_k(x^{(k)} - x^{(k-1)}). \tag{1.3}$$

The scheme (1.3) was first introduced for gradient methods, i.e., when $g \equiv 0$ in problem (1.1), in a seminal work by Nesterov [27], and it was then extended to FB methods by Beck and Teboulle in [5],

[*] The authors are members of the INDAM research group GNCS. This work has been partially supported by MIUR under the project FIRB - Futuro in Ricerca 2012, contract RBFR12M3AC, and by the INDAM research group GNCS.

[†] Dipartimento di Fisica, Matematica e Informatica, Università di Modena e Reggio Emilia (silvia.bonettini@unimore.it)

[‡] Dipartimento di Matematica e di Informatica, Università di Ferrara (simone.rebegoldi@unife.it, valeria.ruggiero@unife.it)

where the authors propose the Fast Iterative Shrinkage–Thresolding Algorithm (FISTA). We remark that slightly different inertial approaches to the one in (1.3) are available in the literature, such as the popular Heavy-ball method, which was formerly proposed for minimizing strongly convex functions with Lipschitz continuous gradient by Polyak in [30], and recently extended to the more general problem (1.1) by Ochs et al. in [28].

The attractiveness of method (1.3) is essentially due to its theoretical convergence properties, as well as its improved numerical performances with respect to standard FB methods. In particular, the most remarkable property of method (1.3) is that it achieves the optimal convergence rate for gradient based methods [5, Theorem 4.4], [27], namely

$$F(x^{(k)}) - F^* = \mathcal{O}\left(\frac{1}{k^2}\right), \tag{1.4}$$

where F^* is the minimum of F. This convergence rate is one order higher than the theoretical $\mathcal{O}(1/k)$ rate typically obtained for standard versions of the FB algorithm. Furthermore, in the very recent paper [1], an improved $o(1/k^2)$ convergence rate result is obtained, mantaining the convergence of the iterates to a solution of problem (1.1), as proved originally in [12] and then, with simplified arguments, in [1]. Finally, the $\mathcal{O}(1/k^2)$ result and the convergence of the iterates of method (1.3) are preserved even when a variable metric is introduced in the proximal gradient step [11], i.e., when the proximal operator in (1.3) is defined with respect to the norm induced by a variable linear operator. This last theoretical result allows to accelerate the practical convergence behaviour of (1.3) by means of a suitable variable metric choice: indeed, the proposed FB method in [11], named Scaled Forward-Backward Extrapolation Method (SFBEM), is shown to outperform the classical FISTA algorithm in several problems arising from signal and image processing.

One of the main limitations of the inertial scheme (1.3) is that it requires the minimum problem (1.2) to be solved exactly at each iteration. This assumption is clearly not realistic in several significant applications where the proximal operator is not available in closed form, such as when g is the Total Variation functional or a more general analysis sparsity prior. In order to overcome this drawback, recent works have provided inexact versions of the inertial FB algorithm [1, 34, 36], in which the proximal gradient step is approximated by applying a finite number of iterations of an optimization solver to the minimum problem in (1.2). However, an unifying convergence analysis of the inexact inertial FB algorithm, where the $o(1/k^2)$ convergence rate and the convergence of the iterates are proved under errors on both the gradient and the proximal step, is still missing. Furthermore, according to the numerical experiences carried out in [34, 36], it emerges that the accelerated FB method (1.3) is more sensitive to computational errors than the standard FB method (1.2) and, consequently, improvements in efficiency might be lost whenever the computational errors do not decay sufficiently fast. This leaves room for additional acceleration techniques aimed at recovering the practical convergence behaviour of (1.3).

In this paper we develop a FB method with extrapolation which generalizes the original FISTA algorithm [5], its inexact versions [1, 34, 36], and the SFBEM method in [11], by introducing simultaneously the following features:

- inexact computation of the proximal operator;
- variable metric;
- adaptive computation of the steplength parameter α_k;
- capability to handle problems where the domain of f is not the whole space $\mathcal{H}$.

Concerning the first feature, the notion of inexactness hereby exploited is based on the ϵ-subdifferential of a convex function, as in [34], and it includes the one proposed in [36] as a special case. A key point of this approach is that the theoretical conditions guaranteeing the convergence can be actually implemented and checked in practice in some cases of interest. Regarding the variable metric, it can be chosen according to any adaptive rule, provided that the sequence of the linear operators inducing the metric converges to a constant operator at a certain rate. Finally, the third feature allows to compute the parameter α_k when the Lipschitz constant is not known, and the fourth one is ensured by simply projecting the extrapolated step on the domain of f whenever it does not belong to it.

The main strength of our proposed algorithm resides, on one hand, in its applicability to a wide class of problems where the proximal operator is not available in closed form and/or the domain of f is not the entire space and, on the other hand, in the possibility of exploiting a variable metric in the proximal gradient step in order to balance the slowdown introduced by computational errors and recover the acceleration typically exhibited by inertial FB methods.

From the theoretical point of view, we prove the $o(1/k^2)$ convergence rate of the objective function values and the convergence of the iterates to a minimizer. Our analysis generalizes (and/or improves) several existing results in the literature. In particular:

- we extend [1, Theorem 1, Theorem 3], which are related to the $o(1/k^2)$ rate and convergence of the iterates respectively, to the case where variable metrics and errors in the gradient and proximal steps are both considered, simplifying the proof of [1, Theorem 3];
- we complete the analysis developed for the inexact FB schemes in [34, 36], where the convergence rate result is only $\mathcal{O}(1/k^2)$, and the convergence of the iterates is not proved;
- we extend the results in [11], concerning the SFBEM algorithm, in presence of errors in the proximal gradient steps and, at the same time, we improve them. Indeed, our $o(1/k^2)$ convergence rate result improves the $\mathcal{O}(1/k^2)$ rate shown in [11, Theorem 12], whereas our convergence result on the iterates removes some unnecessary assumptions on the variable metrics required in [11, Theorem 17].

From the numerical point of view, we report the results obtained on a set of Total Variation based image restoration problems, where we investigate the combined effects of inexactness and variable metric techniques. The results show that the proposed method greatly benefits from the adoption of a variable metric, while being highly competitive with respect to the state-of-the-art algorithms for convex optimization.

The paper is organized as follows. After stating our assumptions and some basic results in Section 2, we present our algorithm in Section 3.1. The convergence properties of the algorithm are then analyzed in Section 3.2: in particular, we prove the $\mathcal{O}(1/k^2)$ convergence rate for the objective function values in Section 3.2.2, we derive the stronger convergence rate $o(1/k^2)$ in Section 3.2.3 and, finally, we give the proof of the weak convergence of the iterates to a minimizer in Section 3.2.4. In Section 4 we discuss in detail the inexactness criterion for the computation of the proximal operator and its practical implementation, while Section 5 describes the numerical results obtained on a TV based image restoration problem. Finally, our conclusions and perspectives are given in Section 6.

2. Problem formulation and preliminaries.

2.1. Notations.

The symbols $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ denote the sets of nonnegative and positive real numbers, respectively. Given an Hilbert space $\mathcal{H}$, we denote with $\|\cdot\|$ the norm induced by the inner product $\langle\cdot,\cdot\rangle$ defined on $\mathcal{H}$. For any function $f:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$, the domain of f is the set $\mathrm{dom}(f)=\{x\in\mathcal{H}:\ f(x)<\infty\}$.

We denote with $\mathcal{S}(\mathcal{H})$ the set of linear, bounded and self-adjoint operators from $\mathcal{H}$ to $\mathcal{H}$. In $\mathcal{S}(\mathcal{H})$, we consider the Loewner partial ordering relation, which is defined as follows:

$$\forall D_1, D_2\in\mathcal{S}(\mathcal{H})\quad D_1\preceq D_2 \Leftrightarrow \langle D_1x,x\rangle\leq\langle D_2x,x\rangle\ \forall x\in\mathcal{H}\,.$$

Let $\mathcal{I}\in\mathcal{S}(\mathcal{H})$ be the identity operator on $\mathcal{H}$. For any $\eta,\gamma\in\mathbb{R}_{>0}$, $\eta\leq\gamma$, we define the following sets

$$\begin{aligned}\mathcal{D}_\eta &= \{D\in S(\mathcal{H}):\ \eta\mathcal{I}\preceq D\}\\ \mathcal{D}_\eta^\gamma &= \{D\in S(\mathcal{H}):\ \eta\mathcal{I}\preceq D\preceq\gamma\mathcal{I}\}.\end{aligned}$$

Clearly we have $\mathcal{D}_\eta^\gamma\subseteq\mathcal{D}_\eta$. If $D\in\mathcal{D}_\eta$, then

$$(x,y):=\langle Dx,y\rangle \tag{2.1}$$

defines an inner product on $\mathcal{H}$. We indicate with $\|\cdot\|_D$ the norm induced by (2.1), i.e., $\|x\|_D^2=\langle Dx,x\rangle$. Then, if $D\in\mathcal{D}_\eta^\gamma$, the following inequality holds

$$\eta\|u\|^2\leq\|u\|_D^2\leq\gamma\|u\|^2,\quad\forall\ u\in\mathcal{H}. \tag{2.2}$$

Remark 2.1. *[16, Theorem 4.6.11] If $D \in \mathcal{D}_\eta$, then D is invertible. If, in addition, $D \in \mathcal{D}_\eta^\gamma$, then we also have $\frac{1}{\gamma}\mathcal{I} \preccurlyeq D^{-1} \preccurlyeq \frac{1}{\eta}\mathcal{I}$.*

Let $Y \subseteq \mathcal{H}$ be a non empty, closed, convex set. The projection operator associated to $D \in \mathcal{D}_\eta$ is defined as

$$P_{Y,D}(x) = \underset{y\in Y}{\operatorname{argmin}} \|y - x\|_D^2, \quad \forall\, x \in \mathcal{H}. \tag{2.3}$$

Remark 2.2. *[11, Lemma 4] The projection operator* (2.3) *is firmly nonexpansive. Therefore, it is in particular nonexpansive, i.e.,*

$$\|P_{Y,D}(x) - P_{Y,D}(y)\|_D \le \|x - y\|_D, \quad \forall\, x, y \in \mathcal{H}.$$

2.2. Problem formulation. In this paper we are interested in solving the optimization problem

$$\min_{x\in\mathcal{H}} F(x) \equiv f(x) + g(x)\,, \tag{2.4}$$

where $\mathcal{H}$ is a Hilbert space, and f and g satisfy the following assumptions:

Assumption 1.

(i) $g : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is a proper, convex, lower semicontinuous function;

(ii) $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is convex and continuously differentiable on a non empty, closed, convex set Y, where dom(g) $\subseteq Y \subseteq$ dom(f);

(iii) f has an $L-$Lipschitz continuous gradient on Y, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in Y.$$

2.3. Basic results of convex analysis. We start by giving the definition of subdifferential of a convex function.

Definition 2.1. *Let $F : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be a convex function. The subdifferential of F at $x \in \mathcal{H}$ is the set*

$$\partial F(x) = \{w \in \mathcal{H} : F(y) \ge F(x) + \langle y - x, w\rangle, \ \ \forall y \in \mathcal{H}\}\,.$$

Remark 2.3. *A point $x \in \mathcal{H}$ is a minimizer of a convex function F if and only if $0 \in \partial F(x)$; then, by observing that the subdifferential of the function F in* (2.4) *is given by $\partial F(x) = \{\nabla f(x)\} + \partial g(x)$ [27, Section 3.1.6], it follows that $x \in \mathcal{H}$ is a solution of problem* (2.4) *if and only if $-\nabla f(x) \in \partial g(x)$.*

Definition 2.2. *The* proximity *or* resolvent *operator associated to a convex function $g : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ in the metric induced by an operator $D \in \mathcal{D}_\eta$ is defined as*

$$\operatorname{prox}_g^D(x) = \arg\min_{z\in\mathcal{H}} g(z) + \frac{1}{2}\|z - x\|_D^2, \quad \forall x \in \mathcal{H}. \tag{2.5}$$

Definition 2.3. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be a continuously differentiable function on the set $Y \subseteq$ dom(f), $g : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ a proper and convex function, $\alpha \in \mathbb{R}_{>0}$ and $D \in \mathcal{D}_\eta$. For any $x \in Y$, the point $p_{\alpha,D}(x)$ is defined as*

$$\begin{aligned} p_{\alpha,D}(x) &= \operatorname{prox}_{\alpha g}^D(x - \alpha D^{-1}\nabla f(x)) \\ &= \underset{y\in\mathcal{H}}{\operatorname{argmin}}\, \mathcal{P}_{\alpha,D}(y; x) \equiv g(y) + \frac{1}{2\alpha}\left\|y - x + \alpha D^{-1}\nabla f(x)\right\|_D^2. \end{aligned}$$

Remark 2.4. *The point $p_{\alpha,D}(x)$ belongs to dom(g), and x is a minimizer of F if and only if $p_{\alpha,D}(x)=x$ [3, Proposition 12.29]. Furthermore, by Definition 2.3 and Remark 2.3, we have*

$$y=p_{\alpha,D}(x) \;\Leftrightarrow\; 0\in\partial\mathcal{P}_{\alpha,D}(y;x) \;\Leftrightarrow\; \frac{1}{\alpha}D\left(x-\alpha D^{-1}\nabla f(x)-y\right)\in\partial g(y). \tag{2.6}$$

When the point $p_{\alpha,D}(x)$ can not be computed explicitly, it is essential to replace it with a computable approximation whose error level can be easily measured. To this purpose, we introduce the concept of ϵ-subdifferential of a convex function.

Definition 2.4. *[38, p. 82] Let $g:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$ be a convex function, $\epsilon\in\mathbb{R}_{\geq0}$. The $\epsilon-$subdifferential of g at $x\in\mathcal{H}$ is the set*

$$\partial_\epsilon g(x)=\{w\in\mathcal{H}: g(y)\geq g(x)+\langle y-x,w\rangle-\epsilon,\quad\forall y\in\mathcal{H}\}.$$

Remark 2.5. *(i) If $\epsilon=0$, then $\partial_\epsilon g(x)=\partial g(x)$, i.e., the ϵ-subdifferential coincides with the exact subdifferential of Definition 2.1.*
(ii) If $x\notin$ dom(g), then $\partial_\epsilon g(x)=\emptyset$ for any $\epsilon\in\mathbb{R}_{\geq0}$. Conversely, if $x\in$ dom(g) and g is lower semicontinuous at x, then $\partial_\epsilon g(x)\neq\emptyset$ for any $\epsilon\in\mathbb{R}_{>0}$ [38, Theorem 2.4.4].

Lemma 2.1. *Let $f,g:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$ be proper, convex, lower semicontinuous functions, $\alpha\in\mathbb{R}_{>0}$ and $\epsilon\in\mathbb{R}_{\geq0}$. Then the following properties hold.*
(i) [38, Theorem 2.4.2] If $g(x)=\alpha f(x)$, then

$$\partial_\epsilon g(x)=\alpha\partial_{\epsilon/\alpha}f(x),\quad\forall\, x\in \text{dom}(f).$$

(ii) [38, Theorem 2.8.7] If there exists $x\in$ dom$(f)\cap$ dom(g) such that g is continuous at x, then

$$\partial_\epsilon(f+g)(x)=\bigcup_{0\leq\epsilon_1+\epsilon_2\leq\epsilon}\partial_{\epsilon_1}f(x)+\partial_{\epsilon_2}g(x).$$

Lemma 2.2. *[21, Example XI 1.2.2] Let $b\in\mathcal{H}$, $D\in\mathcal{D}_\eta$ and define $g(y)=\frac{1}{2}\langle Dy,y\rangle+\langle b,y\rangle$, $\forall\, y\in\mathcal{H}$. Given $\epsilon\in\mathbb{R}_{\geq0}$, we have*

$$\partial_\epsilon g(y)=\{\nabla g(y)\}+\left\{e\in\mathcal{H}:\ \frac{\|e\|^2_{D^{-1}}}{2}\leq\epsilon\right\},\quad\forall\, y\in\mathcal{H}.$$

The notion of ϵ-subdifferential, combined with Remark 2.4, allows us to relax the definition of the proximal gradient point $p_{\alpha,D}(x)$ in the following way.

Definition 2.5. *Let $f:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$ be a continuously differentiable function on the set $Y\subseteq$ dom(f), $g:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$ a proper and convex function, $\alpha\in\mathbb{R}_{>0}$, $D\in\mathcal{D}_\eta$ and $\epsilon\in\mathbb{R}_{\geq0}$. Given $x\in Y$, an ϵ-approximation of the point $y=p_{\alpha,D}(x)$ is any point $\tilde{y}\in\mathcal{H}$ such that*

$$0\in\partial_\epsilon\mathcal{P}_{\alpha,D}(\tilde{y};x)\quad\Leftrightarrow\quad\mathcal{P}_{\alpha,D}(\tilde{y};x)\leq\mathcal{P}_{\alpha,D}(y;x)+\epsilon. \tag{2.7}$$

In this case, we write $\tilde{y}\approx_\epsilon y$.

Remark 2.6. *Since $y\in$ dom(g), then any point $\tilde{y}$ satisfying (2.7) is such that $\tilde{y}\in$ dom(g).*

Remark 2.7. *By applying Lemma 2.1 and Lemma 2.2 to (2.7), it follows that a point $\tilde{y}$ is an ϵ-approximation of $p_{\alpha,D}(x)$ if and only if there exist $w,e\in\mathcal{H}$, $\bar{\epsilon},\hat{\epsilon}\in\mathbb{R}_{\geq0}$ with $\bar{\epsilon}+\hat{\epsilon}\leq\epsilon$, such that the following relation holds:*

$$\nabla f(x)+w+\frac{1}{\alpha}D(\tilde{y}-x+e)=0,\quad w\in\partial_{\bar{\epsilon}}g(\tilde{y}),\ \|e\|^2_D\leq2\alpha\hat{\epsilon} \tag{2.8}$$

or, equivalently,

$$\frac{1}{\alpha}D\left(x-\alpha D^{-1}\nabla f(x)-e-\tilde{y}\right)\in\partial_{\bar{\epsilon}}g(\tilde{y}),\quad \|e\|_D^2\le 2\alpha\hat{\epsilon}. \tag{2.9}$$

Comparing (2.9) *with* (2.6)*, it follows that the parameter* $\bar{\epsilon}$ *controls the error level on the proximal operator, whereas* $\sqrt{\hat{\epsilon}}$ *controls the error level on the gradient step. Definition 2.5 has already been considered elsewhere in the context of proximal-gradient methods. Indeed it is an extension of the so-called "type 1 approximation" treated in [33, Definition 1] and also used in [34], both of which only consider the case* $D\equiv\mathcal{I}$*. In [36, 10], the authors use* (2.9) *with* $\hat{\epsilon}=0$*, which implies* $e=0$ *and no errors in the calculation of the gradient step. Other works have also treated the case* $\bar{\epsilon}=0$*, in which only errors on the gradient are considered (see [33, 1] and references therein). By contrast, our theoretical analysis covers the most general case where both* $\bar{\epsilon}$ *and* $\hat{\epsilon}$ *can be strictly positive and* D *is any positive definite operator.*

We now provide a technical descent lemma which holds for any ϵ-approximation of type (2.7). To this aim, we first recall a slight variant of the well-known descent lemma for functions having a Lipschitz continuous gradient [8, Lemma 6.9.1].

LEMMA 2.3. *[11, Lemma 6] Let* $f:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$ *be a continuously differentiable function with L-Lipschitz continuous gradient on a set* $Y\subseteq dom(f)$*,* $\alpha\in\mathbb{R}_{>0}$*,* $D\in\mathcal{D}_\eta$*. If* $\alpha\le\eta/L$*, we have*

$$f(y)\le f(x)+\langle\nabla f(x),y-x\rangle+\frac{1}{2\alpha}\|y-x\|_D^2,\quad \forall\ x,y\in Y. \tag{2.10}$$

LEMMA 2.4. *Suppose that* $F:\mathcal{H}\to\mathbb{R}\cup\{\infty\}$ *is defined as* $F(x)\equiv f(x)+g(x)$*, with* f *and* g *satisfying Assumption 1. Let* $\alpha\in\mathbb{R}_{>0}$*,* $D\in\mathcal{D}_\eta$*,* $\epsilon\in\mathbb{R}_{\ge0}$*,* $x\in Y$ *and* $\tilde{y}\approx_\epsilon p_{\alpha,D}(x)$*. If* α *is such that* (2.10) *holds, there exist* $\bar{\epsilon},\hat{\epsilon}\in\mathbb{R}_{\ge0}$ *with* $\bar{\epsilon}+\hat{\epsilon}\le\epsilon$ *such that*

$$F(\tilde{y})+\frac{1}{2\alpha}\|z-\tilde{y}\|_D^2\le F(z)+\frac{1}{2\alpha}\|z-x\|_D^2+\bar{\epsilon}+\frac{\sqrt{2\alpha\hat{\epsilon}}}{\alpha}\|z-\tilde{y}\|_D,\quad \forall\ z\in\mathcal{H}. \tag{2.11}$$

Proof. From (2.8), there exist $\bar{\epsilon}\le\epsilon$ and $w\in\mathcal{H}$ such that $w\in\partial_{\bar{\epsilon}}g(\tilde{y})$. By Definition 2.4, this is equivalent to the inequality

$$g(z)\ge g(\tilde{y})+\langle z-\tilde{y},w\rangle-\bar{\epsilon},\quad \forall\ z\in\mathcal{H}. \tag{2.12}$$

Then, the following chain of inequalities holds:

$$\begin{aligned}
F(z) &\ge f(x)+\langle z-x,\nabla f(x)\rangle+g(z)\\
&\ge f(x)+\langle\tilde{y}-x,\nabla f(x)\rangle+g(\tilde{y})+\langle z-\tilde{y},\nabla f(x)+w\rangle-\bar{\epsilon}\\
&= f(x)+\langle\tilde{y}-x,\nabla f(x)\rangle+g(\tilde{y})+\frac{1}{\alpha}\langle z-\tilde{y},D(x-\tilde{y}-e)\rangle-\bar{\epsilon}\\
&= f(x)+\langle\tilde{y}-x,\nabla f(x)\rangle+g(\tilde{y})+\frac{1}{\alpha}\langle z-\tilde{y},D(x-\tilde{y})\rangle-\bar{\epsilon}-\frac{1}{\alpha}\langle z-\tilde{y},De\rangle\\
&= f(x)+\langle\tilde{y}-x,\nabla f(x)\rangle+g(\tilde{y})+\frac{1}{2\alpha}\|x-\tilde{y}\|_D^2+\frac{1}{2\alpha}\|z-\tilde{y}\|_D^2-\frac{1}{2\alpha}\|z-x\|_D^2-\bar{\epsilon}-\frac{1}{\alpha}\langle z-\tilde{y},De\rangle\\
&\ge F(\tilde{y})+\frac{1}{2\alpha}\|z-\tilde{y}\|_D^2-\frac{1}{2\alpha}\|z-x\|_D^2-\bar{\epsilon}-\frac{1}{\alpha}\langle z-\tilde{y},De\rangle.
\end{aligned}$$

where the first inequality follows from the convexity of f, the second one from (2.12), the third equality from the definition of w in (2.8), the fifth one from the basic norm equality

$$\|a-b\|_D^2+\|b-c\|_D^2-\|a-c\|_D^2=2\langle c-b,D(a-b)\rangle \tag{2.13}$$

applied with $a=x$, $b=\tilde{y}$, $c=z$, and the sixth one from the descent condition (2.10). The thesis now follows by applying the Cauchy-Schwarz inequality and observing that $\|e\|_D\le\sqrt{2\alpha\hat{\epsilon}}$. □

3. Algorithm and convergence analysis.

3.1. The proposed algorithm: iSFBEM. In this section we describe in detail the proposed method, denominated inexact Scaled Forward–Backward Extrapolation Method (iSFBEM), which is reported in Algorithm 1.

Algorithm 1 inexact Scaled Forward–Backward Extrapolation Method (iSFBEM)

```
Choose $\alpha_{-1} \in \mathbb{R}_{>0}$, $\eta \in \mathbb{R}_{>0}$, $\delta < 1$, $x^{(-1)} \in Y$ and set $x^{(0)} = x^{(-1)}$. Define two sequences
$\{\beta_k\}_{k\in\mathbb{N}}, \{\epsilon_k\}_{k\in\mathbb{N}}$ with $\beta_k, \epsilon_k \in \mathbb{R}_{\geq 0}$, and a sequence of operators $\{D_k\}_{k\in\mathbb{N}}$ with $D_k \in \mathcal{D}_\eta$.
FOR $k = 0, 1, 2, ...$
    STEP 1. Compute $y^{(k)} = P_{Y,D_k}(x^{(k)} + \beta_k(x^{(k)} - x^{(k-1)}))$.
    STEP 2. Set $\alpha_k = \alpha_{k-1}$, $i_k = 0$.
    STEP 3. Set $\tilde{x}_+^{(k)} \approx_{\epsilon_k} p_{\alpha_k,D_k}(y^{(k)})$.
    STEP 4. If
            (3.1)  $f(\tilde{x}_+^{(k)}) \leq f(y^{(k)}) + \langle \nabla f(y^{(k)}), \tilde{x}_+^{(k)} - y^{(k)} \rangle + \frac{1}{2\alpha_k}\|y^{(k)} - \tilde{x}_+^{(k)}\|_{D_k}^2$
            go to Step 5.
            else set
                $i_k \leftarrow i_k + 1 \quad \alpha_k = \delta^{i_k}\alpha_{k-1}$
            and go to Step 3.
    STEP 5. Set the new iterate $x^{(k+1)} = \tilde{x}_+^{(k)}$.
END
```

Algorithm 1 is an inertial variable metric FB method in which, at each iteration, the following steps are performed:

- the extrapolated–projected point $y^{(k)}$ is computed by projecting $x^{(k)} + \beta_k(x^{(k)} - x^{(k-1)})$ onto the set Y with respect to the metric induced by D_k (Step 1);
- the proximal–gradient point $\tilde{x}_+^{(k)} \approx_{\epsilon_k} p_{\alpha_k,D_k}(y^{(k)})$ is computed inexactly, according to Definition 2.5, using α_k as the steplength parameter and D_k as the metric defining the proximal operator (Steps 2–3);
- the steplength α_k is possibly reduced, according to a backtracking procedure, and Step 3 is repeated, until the descent condition (3.1) is satisfied (Step 4);
- the new iterate $x^{(k+1)}$ is set as the ϵ_k-approximation $\tilde{x}_+^{(k)}$ with α_k computed as in Step 4.

Let us observe that Algorithm 1 is well–defined, i.e., the backtracking loop at Steps 3–4 terminates in a finite number of steps for all $k \geq 0$. Indeed, thanks to Assumption 1(iii), Lemma 2.3 and the fact that $y^{(k)}, \tilde{x}_+^{(k)} \in Y$, condition (3.1) holds for any $\alpha_k \leq \eta/L$, namely for α_k sufficiently small. Furthermore, the sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ is bounded since, by observing that $\{\alpha_k\}_{k\in\mathbb{N}}$ is non–increasing and that the reducing factor is $\delta < 1$, the following inequalities hold

$$0 < \frac{\delta\eta}{L} \leq \alpha_k \leq \alpha_{k-1} \leq \alpha_{-1}. \tag{3.2}$$

Algorithm 1 can be considered as an inexact version of the SFBEM algorithm proposed in [11], and as a generalization of the Fast Iterative Soft Tresholding Algorithm (FISTA) [5]. In particular, Algorithm 1 coincides with the original version of FISTA when $Y = \mathbb{R}^n$, $D_k = \mathcal{I}$, $\epsilon_k = 0$ and $\beta_k = (k-1)/(k+2)$ for all $k \geq 0$. Unlike FISTA, Algorithm 1 allows to handle problems where the proximal operator of the function g does not have a closed formula, and Y does not coincide with the entire space $\mathcal{H}$ and, unlike other inexact versions of FISTA [1, 34, 36], it makes the use of variable metrics in order to speed up its practical convergence rate. We remark that, though not explicitly stated in Algorithm 1, the sequences

$\{\beta_k\}_{k\in\mathbb{N}}$, $\{\epsilon_k\}_{k\in\mathbb{N}}$ and $\{D_k\}_{k\in\mathbb{N}}$ need to be appropriately chosen to make the scheme convergent, as we will see in the subsequent convergence analysis.

3.2. Convergence analysis. In this section, we investigate the convergence rate of the objective function values sequence generated by Algorithm 1 towards the optimal value, as well as the weak convergence of the iterates to a solution of problem (2.4). From now on, we will indicate with x^* any of the solutions of problem (2.4) under Assumption 1 (given in section 2.2), while $\{x^{(k)}\}_{k\in\mathbb{N}}$ will denote the sequence generated by Algorithm 1. We will assume that the parameters sequence $\{\beta_k\}_{k\in\mathbb{N}}$ has the form

$$\beta_k = \begin{cases} 0 & k=0 \\ \frac{\theta_k(1-\theta_{k-1})}{\theta_{k-1}} & k\geq 1 \end{cases} \tag{3.3}$$

where the sequence $\{\theta_k\}_{k\in\mathbb{N}} \subseteq (0,1]$ satisfies

$$\frac{1-\theta_k}{\theta_k^2} \leq \frac{1}{\theta_{k-1}^2}, \quad \forall\, k\geq 0. \tag{3.4}$$

Furthermore, we will use the following notations

$$\begin{aligned} v_k &= F(x^{(k)}) - F(x^*) \\ z^{(k)} &= x^{(k)} + \frac{1-\theta_{k-1}}{\theta_{k-1}}(x^{(k)}-x^{(k-1)}) = x^{(k-1)} + \frac{1}{\theta_{k-1}}(x^{(k)}-x^{(k-1)}) \\ u^{(k)} &= z^{(k)} - x^* \\ t_k &= \frac{1}{\theta_k}. \end{aligned} \tag{3.5}$$

Finally, we will denote with $\bar{\epsilon}_k, \hat{\epsilon}_k$ the error parameters for which Lemma 2.4 holds with $\alpha=\alpha_k$, $D=D_k$, $\epsilon=\epsilon_k$, $x=y^{(k)}$ and $\tilde{y}=x^{(k+1)}$.

3.2.1. Some fundamental lemmas. We now introduce some technical lemmas on which the subsequent convergence analysis relies on. The first one contains a key inequality which generalizes [5, Lemma 4.1] inherent to FISTA, and [11, Lemma 9] related to the SFBEM algorithm, to the case when the proximal gradient step is computed inexactly.

LEMMA 3.1. *Let Assumption 1 hold and suppose that $\{\theta_k\}_{k\in\mathbb{N}}$ satisfies* (3.3)-(3.4). *Then, for all $k\geq 0$, we have*

$$2\alpha_{k+1}t_k^2 v_{k+1} + \|u^{(k+1)}\|_{D_k}^2 \leq 2\alpha_k t_{k-1}^2 v_k + \|u^{(k)}\|_{D_k}^2 + 2\alpha_k t_k^2\bar{\epsilon}_k + 2\sqrt{2\alpha_k\hat{\epsilon}_k}t_k\|u^{(k+1)}\|_{D_k}. \tag{3.6}$$

Proof. The proof follows by proceeding as in [11, Lemma 9]. We define the point $y^*=(1-\theta_k)x^{(k)}+\theta_k x^*$ and observe that $y^*\in \mathrm{dom}(g)$. From (2.11) in Lemma 2.4 with $\tilde{y}=x^{(k+1)}$, $x=y^{(k)}$ and $z=y^*$, we have

$$F(x^{(k+1)}) + \frac{1}{2\alpha_k}\|y^*-x^{(k+1)}\|_{D_k}^2 \leq F(y^*) + \frac{1}{2\alpha_k}\|y^*-y^{(k)}\|_{D_k}^2 + \bar{\epsilon}_k + \frac{\sqrt{2\alpha_k\hat{\epsilon}_k}}{\alpha_k}\|y^*-x^{(k+1)}\|_{D_k}. \tag{3.7}$$

Using the definition of y^* and the convexity of F, we have $F(y^*)\leq (1-\theta_k)F(x^{(k)})+\theta_k F(x^*)$, and from Remark 2.2 it follows that $\|y^*-y^{(k)}\|_{D_k}\leq \|y^*-(x^{(k)}+\beta_k(x^{(k)}-x^{(k-1)}))\|_{D_k}^2$. Furthermore, from the definition of y^* and conditions (3.3)-(3.5), we also have

$$\begin{aligned} \|y^*-x^{(k+1)}\|_{D_k} &= \theta_k\|z^{(k+1)}-x^*\|_{D_k} \\ \|y^*-(x^{(k)}+\beta_k(x^{(k)}-x^{(k-1)}))\|_{D_k} &= \theta_k\|z^{(k)}-x^*\|_{D_k}. \end{aligned}$$

Plugging the previous facts into (3.7), we obtain

$$F(x^{(k+1)}) + \frac{\theta_k^2}{2\alpha_k}\|z^{(k+1)} - x^*\|_{D_k}^2 \le (1-\theta_k)F(x^{(k)}) + \theta_k F(x^*) + \frac{\theta_k^2}{2\alpha_k}\|z^{(k)} - x^*\|_{D_k}^2 + $$
$$+ \bar{\epsilon}_k + \frac{\theta_k}{\alpha_k}\sqrt{2\alpha_k\hat{\epsilon}_k}\|z^{(k+1)} - x^*\|_{D_k}.$$

Subtracting $F(x^*)$ from both sides, multiplying both sides by $2\alpha_k/\theta_k^2$ and rearranging terms gives

$$2\frac{\alpha_k}{\theta_k^2}v_{k+1} + \|z^{(k+1)} - x^*\|_{D_k}^2 \le 2\alpha_k\frac{1-\theta_k}{\theta_k^2}v_k + \|z^{(k)} - x^*\|_{D_k}^2 + \frac{2\alpha_k}{\theta_k^2}\bar{\epsilon}_k + \frac{2}{\theta_k}\sqrt{2\alpha_k\hat{\epsilon}_k}\|z^{(k+1)} - x^*\|_{D_k}. \tag{3.8}$$

Finally, observing that $\alpha_{k+1} \le \alpha_k$, we obtain

$$2\frac{\alpha_{k+1}}{\theta_k^2}v_{k+1} + \|z^{(k+1)} - x^*\|_{D_k}^2 \le 2\alpha_k\frac{1-\theta_k}{\theta_k^2}v_k + \|z^{(k)} - x^*\|_{D_k}^2 + \frac{2\alpha_k}{\theta_k^2}\bar{\epsilon}_k + \frac{2}{\theta_k}\sqrt{2\alpha_k\hat{\epsilon}_k}\|z^{(k+1)} - x^*\|_{D_k}. \tag{3.9}$$

The thesis now follows by applying (3.4) to the previous inequality. □

We also recall the following two lemmas which hold for nonnegative sequences.

Lemma 3.2. *[31] Let $\{p_k\}_{k\in\mathbb{N}}$, $\{\zeta_k\}_{k\in\mathbb{N}}$ and $\{\xi_k\}_{k\in\mathbb{N}}$ be sequences of real nonnegative numbers such that $p_{k+1} \le (1+\zeta_k)p_k + \xi_k$ and $\sum_{k=0}^\infty \zeta_k < \infty$, $\sum_{k=0}^\infty \xi_k < \infty$. Then, $\{p_k\}_{k\in\mathbb{N}}$ converges.*

Lemma 3.3. *[34, Lemma 1] Let $\{p_k\}_{k\in\mathbb{N}}$, $\{q_k\}_{k\in\mathbb{N}}$, $\{\lambda_k\}_{k\in\mathbb{N}}$ be sequences of real nonnegative numbers, with $\{q_k\}_{k\in\mathbb{N}}$ being a monotone nondecreasing sequence, satisfying the following recursive property*

$$p_k^2 \le q_k + \sum_{i=1}^k \lambda_i p_i, \quad \forall\ k \ge 1. \tag{3.10}$$

Then the following inequality holds:

$$p_k \le \frac{1}{2}\sum_{i=1}^k \lambda_i + \left(q_k + \left(\frac{1}{2}\sum_{i=1}^k \lambda_i\right)^2\right)^{\frac{1}{2}}, \quad \forall\ k \ge 1. \tag{3.11}$$

3.2.2. An $\mathcal{O}(1/k^2)$ convergence rate result. Our aim is now to show that the sequence of the function values $\{v_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 with β_k chosen as in (3.13) with $a \ge 2$ has an $\mathcal{O}(1/k^2)$ convergence rate. More precisely, the convergence rate result will be proved under the following assumptions on the parameters of Algorithm 1.

Assumption 2. *The sequence of operators $\{D_k\}_{k\in\mathbb{N}} \subseteq \mathcal{D}_\eta$ satisfies*

$$D_{k+1} \preceq (1+\eta_k)D_k \quad \forall\ k \ge 0 \quad \textit{with } \eta_k \in \mathbb{R}, \eta_k \ge 0 \textit{ such that } \sum_{k=0}^\infty \eta_k < \infty. \tag{3.12}$$

Assumption 3. *Given $a \ge 2$, the sequences $\{\beta_k\}_{k\in\mathbb{N}}$, $\{\theta_k\}_{k\in\mathbb{N}}$ are computed as*

$$\theta_k = \begin{cases} 1 & k = -1, 0 \\ \frac{a}{k+a} & k \ge 1 \end{cases} \qquad \beta_k = \begin{cases} 0 & k = 0 \\ \frac{k-1}{k+a} & k \ge 1 \end{cases} \tag{3.13}$$

Assumption 4. *The sequences $\{k^2\bar{\epsilon}_k\}_{k\in\mathbb{N}}$ and $\{k\sqrt{\hat{\epsilon}_k}\}_{k\in\mathbb{N}}$ are summable.*

Remark 3.1. *Variable metrics satisfying condition (3.12) were first considered in the context of iterative methods in [15]. As proved in [15, Lemma 2.3], if $\{D_k\}_{k\in\mathbb{N}} \subseteq D_\eta^\gamma$ is a sequence of linear operators satisfying condition (3.12), then there exists $D \in \mathcal{D}_\eta$ such that D_k converges to D pointwise.*

Remark 3.2. *It is easy to see that the sequences $\{\beta_k\}_{k\in\mathbb{N}}$, $\{\theta_k\}_{k\in\mathbb{N}}$ computed as in* (3.13) *satisfy* (3.3)-(3.4) *[11, p. A2565]. This choice for the extrapolation parameters $\{\beta_k\}_{k\in\mathbb{N}}$ has been first proposed for the FISTA algorithm, with $a=2$, in [5], and then generalized to the case $a\geq 2$ in [12].*

Remark 3.3. *The choice $\epsilon_k=\mathcal{O}(1/k^p)$, with $p>4$, is sufficient in general to guarantee Assumption 4. Furthermore, if $\hat{\epsilon}_k\equiv 0$, i.e., if there are no errors in the computation of the gradient, we can relax the previous sufficient condition and take $\epsilon_k=\mathcal{O}(1/k^p)$, with $p>3$.*

The convergence rate result will now be proven by combining arguments derived from [34, Proposition 2] and [11, Theorem 12]. On one hand, our result extends the one in [34] by taking into account the presence of the projection onto Y in the extrapolation step, the positive definite operator D_k in the definition of the approximate proximal point $\tilde{x}_+^{(k)}$ and the backtracking procedure to compute α_k. On the other hand, our result requires the same crucial assumption (3.12) on the sequence of operators $\{D_k\}_{k\in\mathbb{N}}$ considered in [11].

Theorem 3.1. *Let Assumptions 1-4 hold. Define the sequences*

$$A_k=\sum_{i=0}^{k}(i+1)\sqrt{2\alpha_0\hat{\epsilon}_i},\quad B_k=\sum_{i=0}^{k}(i+1)^2\alpha_0\bar{\epsilon}_i,\quad C_k=\prod_{i=0}^{k-1}(1+\eta_i).$$

Then $A=\lim_{k\to+\infty}A_k\in\mathbb{R}_{\geq 0}$, $B=\lim_{k\to+\infty}B_k\in\mathbb{R}_{\geq 0}$, $C=\lim_{k\to+\infty}C_k\in\mathbb{R}_{>0}$ and, for all $k\geq 1$, the following bound on the function values holds:

$$F(x^{(k+1)})-F(x^*)\leq\frac{CLa^2\left(\|x^{(0)}-x^*\|_{D_0}+2\sqrt{C}A+\sqrt{2B}\right)^2}{2\delta\eta(k+a)^2}.\tag{3.14}$$

Proof. By Assumption 4, we have $A,B\in\mathbb{R}_{\geq 0}$ and, by Lemma 3.2, also $C\in\mathbb{R}_{>0}$. Setting $s_k=2\alpha_k t_{k-1}^2 v_k$ and applying recursively (3.6) and (3.12), we obtain

$$\begin{aligned}
s_{k+1} &+ \|u^{(k+1)}\|_{D_k}^2\leq\\
&\overset{(3.6)}{\leq} s_k+\|u^{(k)}\|_{D_k}^2+\frac{2\alpha_k}{\theta_k^2}\bar{\epsilon}_k+\frac{2\sqrt{2\alpha_k\hat{\epsilon}_k}}{\theta_k}\|u^{(k+1)}\|_{D_k}\\
&\overset{(3.12)}{\leq}(1+\eta_{k-1})(s_k+\|u^{(k)}\|_{D_{k-1}}^2)+\frac{2\alpha_k}{\theta_k^2}\bar{\epsilon}_k+\frac{2\sqrt{2\alpha_k\hat{\epsilon}_k}}{\theta_k}\|u^{(k+1)}\|_{D_k}\\
&\overset{(3.6)}{\leq}(1+\eta_{k-1})\left(s_{k-1}+\|u^{(k-1)}\|_{D_{k-1}}^2+\frac{2\alpha_{k-1}}{\theta_{k-1}^2}\bar{\epsilon}_{k-1}+\frac{2\sqrt{2\alpha_{k-1}\hat{\epsilon}_{k-1}}}{\theta_{k-1}}\|u^{(k)}\|_{D_{k-1}}\right)+\\
&+\frac{2\alpha_k}{\theta_k^2}\bar{\epsilon}_k+\frac{2\sqrt{2\alpha_k\hat{\epsilon}_k}}{\theta_k}\|u^{(k+1)}\|_{D_k}\\
&\vdots\\
&\leq\left(\prod_{i=1}^{k-1}(1+\eta_i)\right)(s_1+\|u^{(1)}\|_{D_1}^2)+2\sum_{i=1}^{k}\left(\prod_{j=i}^{k-1}(1+\eta_j)\right)\left(\frac{\alpha_i}{\theta_i^2}\bar{\epsilon}_i+\frac{\sqrt{2\alpha_i\hat{\epsilon}_i}}{\theta_i}\|u^{(i+1)}\|_{D_i}\right).
\end{aligned}\tag{3.15}$$

where $\Pi_{j=i}^{k-1}(1+\eta_j)=1$ when $i>k-1$. If we apply one last time (3.12) and then use (3.9) instead of (3.6), we get

$$\begin{aligned}
s_{k+1}+\|u^{(k+1)}\|_{D_k}^2\leq&\left(\prod_{i=0}^{k-1}(1+\eta_i)\right)\left(2\alpha_0\frac{1-\theta_0}{\theta_0}v_0+\|u^{(0)}\|_{D_0}^2\right)+\\
&+2\sum_{i=0}^{k}\left(\prod_{j=i}^{k-1}(1+\eta_j)\right)\left(\frac{\alpha_i}{\theta_i^2}\bar{\epsilon}_i+\frac{\sqrt{2\alpha_i\hat{\epsilon}_i}}{\theta_i}\|u^{(i+1)}\|_{D_i}\right).
\end{aligned}$$

Since $\theta_0 = 1$, $C_k = \prod_{i=0}^{k-1}(1+\eta_i) = (1+\eta_{k-1})C_{k-1}$ is a convergent sequence (see Lemma 3.2) and $\prod_{j=i}^{k-1}(1+\eta_j) \leq C_k \leq C$, the previous inequality yields

$$s_{k+1} + \|u^{(k+1)}\|_{D_k}^2 \leq C\left(\|u^{(0)}\|_{D_0}^2 + 2\sum_{i=0}^{k}\frac{\alpha_i}{\theta_i^2}\bar{\epsilon}_i\right) + \sum_{i=0}^{k}\frac{2C\sqrt{2\alpha_i\hat{\epsilon}_i}}{\theta_i}\|u^{(i+1)}\|_{D_i}, \quad \forall\, k \geq 0. \tag{3.16}$$

Proceeding as in [34, Proposition 2], we now exploit (3.16) to bound first the quantity $\|u^{(i+1)}\|_{D_i}$ and then the function values. Discarding the nonnegative quantity s_{k+1} from the left-hand side of (3.16) and recalling the upper bound in (3.2), we can use Lemma 3.3 with $p_k = \|u^{(k)}\|_{D_{k-1}}$, $q_k = C\left(\|u^{(0)}\|_{D_0}^2 + 2\sum_{i=0}^{k-1}\frac{\alpha_0}{\theta_i^2}\bar{\epsilon}_i\right)$, and $\lambda_k = \frac{2C\sqrt{2\alpha_0\hat{\epsilon}_{k-1}}}{\theta_{k-1}}$ and obtain

$$\|u^{(k+1)}\|_{D_k} \leq C\sum_{i=0}^{k}\frac{\sqrt{2\alpha_0\hat{\epsilon}_i}}{\theta_i} + \left(C\left(\|u^{(0)}\|_{D_0}^2 + 2\sum_{i=0}^{k}\frac{\alpha_0\bar{\epsilon}_i}{\theta_i^2}\right) + \left(C\sum_{i=0}^{k}\frac{\sqrt{2\alpha_0\hat{\epsilon}_i}}{\theta_i}\right)^2\right)^{\frac{1}{2}}. \tag{3.17}$$

By definition of θ_k given in (3.13), we have $1/\theta_i \leq i+1$, for all $i \geq 0$. Therefore, equation (3.17) can be rewritten in terms of the sequences A_k and B_k as

$$\|u^{(k+1)}\|_{D_k} \leq CA_k + \sqrt{C}\left(\|u^{(0)}\|_{D_0}^2 + 2B_k + CA_k^2\right)^{\frac{1}{2}}. \tag{3.18}$$

In general, for $i = 0, \dots, k$, we can bound the quantity $\|u^{(i+1)}\|_{D_i}$ in the following way:

$$\begin{aligned}\|u^{(i+1)}\|_{D_i} &\leq CA_i + \sqrt{C}\left(\|u^{(0)}\|_{D_0}^2 + 2B_i + CA_i^2\right)^{\frac{1}{2}} \leq \sqrt{C}\|u^{(0)}\|_{D_0} + 2CA_i + \sqrt{2CB_i}\\ &\leq \sqrt{C}\|u^{(0)}\|_{D_0} + 2CA_k + \sqrt{2CB_k}\end{aligned} \tag{3.19}$$

where the third inequality follows from the fact that $\{A_k\}_{k\in\mathbb{N}}$ and $\{B_k\}_{k\in\mathbb{N}}$ are nondecreasing sequences. Going back to equation (3.16), we will now show that the function values are bounded. Discarding $\|u^{(k+1)}\|_{D_k}^2$ from the left-hand side of (3.16), recalling the definition of s_k, the upper bound in (3.2), formula (3.13) which implies $1/\theta_i \leq i+1$, and applying (3.19) to the right-hand side of (3.16), we get

$$\begin{aligned}\frac{2\alpha_{k+1}}{\theta_k^2}\left(F(x^{(k+1)}) - F(x^*)\right) &\leq C\left(\|x^{(0)} - x^*\|_{D_0}^2 + 2B_k + 2A_k\left(\sqrt{C}\|x^{(0)} - x^*\|_{D_0} + 2CA_k + \sqrt{2CB_k}\right)\right)\\ &= C\left(\|x^{(0)} - x^*\|_{D_0}^2 + 2\sqrt{C}A_k\|x^{(0)} - x^*\|_{D_0} + 4CA_k^2 + 2B_k + 2\sqrt{C}A_k\sqrt{2B_k}\right)\\ &\leq C\left(\|x^{(0)} - x^*\|_{D_0} + 2\sqrt{C}A_k + \sqrt{2B_k}\right)^2.\end{aligned}$$

The thesis follows from the lower bound in (3.2), formula (3.13), which implies $1/\theta_k^2 = (k+a)^2/a^2$, and the monotonicity of the sequences $\{A_k\}_{k\in\mathbb{N}}$ and $\{B_k\}_{k\in\mathbb{N}}$. □

By combining Theorem 3.1 with Remark 3.3, we are also able to state the following corollary.

Corollary 3.1. *Let Assumptions 1-3 hold and suppose that one of the following conditions is satisfied:*

- $\epsilon_k = \mathcal{O}(1/k^p)$, *with* $p > 4$;
- $\hat{\epsilon}_k \equiv 0$ *and* $\epsilon_k = \bar{\epsilon}_k = \mathcal{O}(1/k^p)$, *with* $p > 3$.

Then we have:

$$F(x^{(k)}) - F(x^*) = \mathcal{O}\left(\frac{1}{k^2}\right).$$

Remark 3.4. *We observe that Theorem 3.1 includes, as special cases, the following results obtained in related previous works:*

- *[11, Theorem 12], when $\epsilon_k \equiv 0$;*
- *[5, Theorem 4.4], which is recovered when $Y = \mathbb{R}^n$, $a = 2$, $D_k = \mathcal{I}$ and $\epsilon_k \equiv 0$;*
- *[34, Proposition 2], when $Y = \mathbb{R}^n$, $a = 2$, $D_k = \mathcal{I}$ and $\alpha_k \equiv 1/L$;*
- *[12, Theorem 3], when $Y = \mathbb{R}^n$, $D_k = \mathcal{I}$, $\alpha_k \equiv 1/L$ and $\epsilon_k \equiv 0$.*
- *[36, Theorem 4.4], when $Y = \mathbb{R}^n$, $D_k = \mathcal{I}$, $\alpha_k \equiv 1/L$ and $\hat{\epsilon}_k \equiv 0$.*

3.2.3. An improved convergence rate result: from $\mathcal{O}(1/k^2)$ to $o(1/k^2)$. In the following, we show that a slightly improved convergence rate result can be obtained for Algorithm 1 if the parameter β_k in (3.13) is chosen with $a > 2$. In this case, we prove that the rate of convergence of the sequence $\{v_k\}_{k\in\mathbb{N}}$ is $o(1/k^2)$, rather than $\mathcal{O}(1/k^2)$. This result extends the one obtained for FISTA in [1, Theorem 1] by taking into account the inexactness of the proximal operator and the presence of a variable metric. We remark that, although the authors in [1] have also extended their convergence result to an inexact version of FISTA, they only consider the case in which the proximal-gradient point is computed according to equation (2.9) with $\bar{\epsilon} = 0$, that is, only the case where the proximal operator is evaluated exactly in a perturbed gradient step, whereas we treat the case where both the gradient and the proximal operator are computed inexactly.

We start by proving the summability of the sequence $\{kv_k\}_{k\in\mathbb{N}}$. This result generalizes the one in [12, Theorem 2], which does not take into account neither the variable metric choice nor the inexactness of the proximal operator, and the one in [11, Lemma 13], where only the variable metric is considered.

Lemma 3.4. *Let Assumptions 1-2-4 hold and suppose that Assumption 3 holds with $a > 2$. Then the sequence $\{kv_k\}_{k\in\mathbb{N}}$ is summable.*

Proof. Observing that $t_k^2 = \frac{1}{\theta_k^2}$, we can write the inequality (3.9) as follows:

$$\alpha_{k+1}t_k^2 v_{k+1} - \alpha_k(t_k^2 - t_k)v_k \le \frac{\|u^{(k)}\|_{D_k}^2}{2} - \frac{\|u^{(k+1)}\|_{D_k}^2}{2} + \alpha_k t_k^2 \bar{\epsilon}_k + t_k\|u^{(k+1)}\|_{D_k}\sqrt{2\alpha_k\hat{\epsilon}_k}\,. \tag{3.20}$$

Equation (3.19) implies that there exists $U \in \mathbb{R}_{\ge 0}$ such that $\|u^{(k+1)}\|_{D_k} \le U$, with $U = \sqrt{C}\|x^{(0)} - x^*\|_{D_0} + 2CA + (2CB)^{1/2}$. Then, proceeding as in [11, Lemma 13], we sum up (3.20) from $k = 0, ..., K$ and, by applying (3.12), recalling the upper bound in (3.2) and the fact that $t_k \le k + 1$, we obtain

$$\alpha_{K+1}t_K^2 v_{K+1} + \sum_{k=1}^{K} \alpha_k(t_{k-1}^2 - t_k^2 + t_k)v_k \le \frac{1}{2}\sum_{k=0}^{K-1}\eta_k\|u^{(k+1)}\|_{D_k}^2 + \frac{\|u^{(0)}\|_{D_0}^2}{2} + B_K + UA_K.$$

Furthermore, using again the upper bound on $\|u^{(k+1)}\|_{D_k}$ and the lower bound in (3.2), we obtain

$$\sum_{k=1}^{K}(t_{k-1}^2 - t_k^2 + t_k)v_k \le \frac{L}{\delta\eta}\left(\frac{U^2}{2}\sum_{k=0}^{K-1}\eta_k + \frac{\|u^{(0)}\|_{D_0}^2}{2} + B_K + UA_K\right).$$

By Remark 3.2 and the fact that η_k, A_k, and B_k are all summable sequences, it follows that $\{(t_{k-1}^2 - t_k^2 + t_k)v_k\}_{k\in\mathbb{N}}$ is a nonnegative summable sequence. Finally, observing that

$$t_{k-1}^2 - t_k^2 + t_k = \frac{k(a-2) + (a-1)^2}{a^2}\,,$$

and recalling that $a > 2$, we can conclude that also $\{kv_k\}_{k\in\mathbb{N}}$ is summable. □

The following Lemma is divided into three parts. In the first one, we show that the sequence $\{k^2\|x^{(k-1)} - x^{(k)}\|_{D_{k-1}}^2\}_{k\in\mathbb{N}}$ is bounded; this is done by following the same technique used in Theorem 3.1. In the second one, taking inspiration from [1, Lemma 2], we prove that $\{k^2\|x^{(k-1)} - x^{(k)}\|_{D_{k-1}}^2 +$

$\alpha_k k^2(F(x^{(k)}) - F(x^*))\}_{k\in\mathbb{N}}$ converges, using the previous point of the Lemma to deal with the inexactness of the gradient step. Finally, the last part is concerned with the summability of the sequence $\{k\|x^{(k)} - x^{(k-1)}\|^2_{D_{k-1}}\}_{k\in\mathbb{N}}$; this statement not only extends the result contained in [11, Lemma 14] to the presence of inexact proximal-gradient steps, but also improves it, since it removes the $\mathcal{O}(1/k^p)$, $p > 2$ convergence rate assumption on the parameters $\{\eta_k\}_{k\in\mathbb{N}}$ which was required instead in [11, Lemma 14].

LEMMA 3.5. *Let Assumptions 1-2-4 hold, suppose that Assumption 3 holds with $a > 2$ and set $\delta_k = \|x^{(k-1)} - x^{(k)}\|^2_{D_{k-1}}/2$. Then the following statements hold:*
(i) $\{k^2\delta_k\}_{k\in\mathbb{N}}$ is a bounded sequence;
(ii) $\{k^2\delta_k + \alpha_k k^2 v_k\}_{k\in\mathbb{N}}$ converges;
(iii) $\{k\delta_k\}_{k\in\mathbb{N}}$ is a summable sequence.

Proof. (i) From (2.11) with $x = y^{(k)}$, $\tilde{y} = x^{(k+1)}$ and $z = x^{(k)}$, it follows that

$$F(x^{(k+1)}) + \frac{\|x^{(k)} - x^{(k+1)}\|^2_{D_k}}{2\alpha_k} \le F(x^{(k)}) + \frac{\|x^{(k)} - y^{(k)}\|^2_{D_k}}{2\alpha_k} + \bar{\epsilon}_k + \frac{\sqrt{2\alpha_k\hat{\epsilon}_k}}{\alpha_k}\|x^{(k)} - x^{(k+1)}\|_{D_k}. \tag{3.21}$$

From Remark 2.2, since $x^{(k)} \in \mathrm{dom}(g) \subseteq Y$, we have

$$\|x^{(k)} - y^{(k)}\|^2_{D_k} \le \beta_k^2\|x^{(k)} - x^{(k-1)}\|^2_{D_k}.$$

Then, subtracting $F(x^*)$ from both sides of (3.21) and using the definition of β_k in (3.13), we can write

$$v_{k+1} + \frac{\|x^{(k)} - x^{(k+1)}\|^2_{D_k}}{2\alpha_k} \le v_k + \left(\frac{k-1}{k+a}\right)^2 \frac{\|x^{(k)} - x^{(k-1)}\|^2_{D_k}}{2\alpha_k} + \bar{\epsilon}_k + \frac{\sqrt{2\alpha_k\hat{\epsilon}_k}}{\alpha_k}\|x^{(k)} - x^{(k+1)}\|_{D_k}. \tag{3.22}$$

Since $k + a \ge k + 1$, (3.22) implies

$$\begin{aligned}\alpha_k(k+1)^2 v_{k+1} + (k+1)^2\frac{\|x^{(k)} - x^{(k+1)}\|^2_{D_k}}{2} &\le \alpha_k(k+1)^2 v_k + (k-1)^2\frac{\|x^{(k-1)} - x^{(k)}\|^2_{D_k}}{2} + \\ &+ \alpha_k(k+1)^2\bar{\epsilon}_k + (k+1)^2\sqrt{2\alpha_k\hat{\epsilon}_k}\|x^{(k)} - x^{(k+1)}\|_{D_k}.\end{aligned} \tag{3.23}$$

For all k we have

$$\alpha_k(k+1)^2 v_k = \alpha_k(k^2 + 1 + 2k)v_k = \alpha_k k^2 v_k + \alpha_k(1 + 2k)v_k.$$

Then using the above equation with (3.2), we can conveniently rewrite (3.23) as

$$\begin{aligned}\alpha_{k+1}(k+1)^2 v_{k+1} + (k+1)^2\frac{\|x^{(k)} - x^{(k+1)}\|^2_{D_k}}{2} &\le \alpha_k k^2 v_k + (k-1)^2\frac{\|x^{(k-1)} - x^{(k)}\|^2_{D_k}}{2} + \alpha_0(k+1)^2\bar{\epsilon}_k + \\ &+ \alpha_0(1+2k)v_k + \sqrt{2}\left((k+1)\sqrt{2\alpha_0\hat{\epsilon}_k}\right)\left((k+1)\frac{\|x^{(k)} - x^{(k+1)}\|_{D_k}}{\sqrt{2}}\right).\end{aligned} \tag{3.24}$$

Using (3.12) in the right-hand side of (3.24) and recalling the definition of δ_k, (3.24) can be rewritten as

$$\begin{aligned}\alpha_{k+1}(k+1)^2 v_{k+1} + (k+1)^2\delta_{k+1} &\le \alpha_k k^2 v_k + (k-1)^2\left(1 + \eta_{k-1}\right)\delta_k + \alpha_0(k+1)^2\bar{\epsilon}_k + \alpha_0(1+2k)v_k + \\ &+ \sqrt{2}\left((k+1)\sqrt{2\alpha_0\hat{\epsilon}_k}\right)\left((k+1)\sqrt{\delta_{k+1}}\right).\end{aligned} \tag{3.25}$$

Observing that $k - 1 \le k$, we can apply recursively (3.12) and (3.25) to the right-hand side of (3.25) for k times, and then discard the nonnegative quantity $\alpha_{k+1}(k+1)^2 v_{k+1}$ from the left-hand side of (3.25), to obtain

$$\begin{aligned}(k+1)^2\delta_{k+1} \le \alpha_0\sum_{i=0}^{k}\left(\prod_{j=i}^{k-1}(1+\eta_j)\right)(i+1)^2\bar{\epsilon}_i + \alpha_0\sum_{i=0}^{k}\left(\prod_{j=i}^{k-1}(1+\eta_j)\right)(1+2i)v_i + \\ + \sqrt{2}\sum_{i=0}^{k}\left(\prod_{j=i}^{k-1}(1+\eta_j)\right)\left((i+1)\sqrt{2\alpha_0\hat{\epsilon}_i}\right)\left((i+1)\sqrt{\delta_{i+1}}\right)\end{aligned}$$

which, by recalling that $\prod_{j=i}^{k-1}(1+\eta_j) \le C_k \le C$, yields

$$(k+1)^2\delta_{k+1} \le C\left(\sum_{i=0}^{k}(i+1)^2\alpha_0\bar{\epsilon}_i + \sum_{i=0}^{k}(1+2i)\alpha_0 v_i\right) + \sum_{i=0}^{k}\left(\sqrt{2}C(i+1)\sqrt{2\alpha_0\hat{\epsilon}_i}\right)\left((i+1)\sqrt{\delta_{i+1}}\right). \tag{3.26}$$

Recalling the definitions of the sequences A_k and B_k, introducing the new sequence

$$E_k = \sum_{i=0}^{k}(1+2i)\alpha_0 v_i,$$

and applying Lemma 3.3 to (3.26), we can bound $(k+1)\sqrt{\delta_{k+1}}$ as follows:

$$(k+1)\sqrt{\delta_{k+1}} \le \frac{CA_k}{\sqrt{2}} + \left(C\left(B_k+E_k\right) + \frac{(CA_k)^2}{2}\right)^{\frac{1}{2}}. \tag{3.27}$$

Thanks to Lemma 3.4, E_k is summable and, since $\{k^2\bar{\epsilon}_k\}_{k\in\mathbb{N}}$ and $\{k\sqrt{\hat{\epsilon}_k}\}_{k\in\mathbb{N}}$ are assumed summable, then also A_k and B_k are summable. Then $\{k^2\delta_k\}_{k\in\mathbb{N}}$ is a bounded sequence.
(ii) Setting $M = C\lim_k A_k + \left(2C(\lim_k B_k + \lim_k E_k) + (C\lim_k A_k)^2\right)^{1/2}$, from (3.27) the following bound holds:

$$\sqrt{2}(k+1)\sqrt{\delta_{k+1}} \le M. \tag{3.28}$$

Applying this bound to (3.25) with $k-1 \le k$, we get

$$\begin{aligned}\alpha_{k+1}(k+1)^2 v_{k+1} + (k+1)^2\delta_{k+1} \le \alpha_k k^2 v_k + k^2(1+\eta_{k-1})\delta_k + \alpha_0(k+1)^2\bar{\epsilon}_k + \alpha_0(1+2k)v_k + \\ + M(k+1)\sqrt{2\alpha_0\hat{\epsilon}_k},\end{aligned} \tag{3.29}$$

which, setting $\zeta_k = \alpha_0(k+1)^2\bar{\epsilon}_k + \alpha_0(1+2k)v_k + M(k+1)\sqrt{2\alpha_0\hat{\epsilon}_k}$, becomes

$$\alpha_{k+1}(k+1)^2 v_{k+1} + (k+1)^2\delta_{k+1} \le (1+\eta_{k-1})\left(\alpha_k k^2 v_k + k^2\delta_k\right) + \zeta_k. \tag{3.30}$$

Since $\{k^2\bar{\epsilon}_k\}_{k\in\mathbb{N}}$, $\{k\sqrt{\hat{\epsilon}_k}\}_{k\in\mathbb{N}}$ and $\{kv_k\}_{k\in\mathbb{N}}$ are summable, then also $\{\zeta_k\}_{k\in\mathbb{N}}$ is a summable sequence. Then (ii) follows from Lemma 3.2.
(iii) Again from (3.28), we have the bound

$$(k-1)^2\delta_k\eta_{k-1} \le k^2\delta_k\eta_{k-1} \le M^2\eta_{k-1}$$

which applied to (3.25) yields

$$\begin{aligned}\alpha_{k+1}(k+1)^2 v_{k+1} + (k+1)^2\delta_{k+1} \le \alpha_k k^2 v_k + (k-1)^2\delta_k + \alpha_0(k+1)^2\bar{\epsilon}_k + \alpha_0(1+2k)v_k + \\ + M(k+1)\sqrt{2\alpha_0\hat{\epsilon}_k} + M^2\eta_{k-1}.\end{aligned}$$

Defining $\overline{\zeta}_k = \zeta_k + M^2\eta_{k-1}$, the previous inequality also reads as

$$\alpha_{k+1}(k+1)^2 v_{k+1} + (k+1)^2\delta_{k+1} \le \alpha_k k^2 v_k + (k-1)^2\delta_k + \overline{\zeta}_k.$$

Observing that $(k+1)^2 \ge k^2+k+1$, we obtain

$$\alpha_{k+1}(k+1)^2 v_{k+1} + k^2\delta_{k+1} + (k+1)\delta_{k+1} \le \alpha_k k^2 v_k + (k-1)^2\delta_k + \overline{\zeta}_k. \tag{3.31}$$

Summing up the previous inequality for $k = 1,\ldots,K$ gives

$$\alpha_{K+1}(K+1)^2 v_{K+1} + K^2\delta_{K+1} + \sum_{k=1}^{K}(k+1)\delta_{k+1} \le \alpha_1 v_1 + \sum_{k=1}^{K}\overline{\zeta}_k.$$

Since $\{\zeta_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ are summable, then also $\{\overline{\zeta}_k\}_{k\in\mathbb{N}}$ is a summable sequence. From this observation, the thesis follows. □

We are now ready to give the $o(1/k^2)$ convergence rate result.

Theorem 3.2. *Let Assumptions 1-2-4 hold and suppose that Assumption 3 holds with $a > 2$. Then*

$$\lim_{k\to+\infty} k^2(F(x^{(k)}) - F(x^*)) = 0, \qquad \lim_{k\to+\infty} k\|x^{(k)} - x^{(k-1)}\| = 0,$$

that is, $F(x^{(k)}) - F(x^*) = o(1/k^2)$ and $\|x^{(k)} - x^{(k-1)}\| = o(1/k)$.

Proof. The proof is similar to the one in [1, Theorem 1]. From Lemma 3.4, point (iii) of Lemma 3.5 and (3.2), we deduce that

$$\sum_{k=1}^{+\infty} \frac{1}{k}\left(k^2\delta_k + \alpha_k k^2 v_k\right) < +\infty.$$

Combining this with point (ii) of Lemma 3.5, it necessarily follows that

$$\lim_{k\to+\infty} k^2\|x^{(k)} - x^{(k-1)}\|^2_{D_{k-1}} + \alpha_k k^2(F(x^{(k)}) - F(x^*)) = 0. \tag{3.32}$$

Furthermore, since $\eta\|x^{(k)} - x^{(k-1)}\|^2 \leq \|x^{(k)} - x^{(k-1)}\|^2_{D_{k-1}}$ and $\alpha_k \geq (\delta\eta)/L$, the previous limit yields the following one

$$\lim_{k\to+\infty} \eta k^2\|x^{(k)} - x^{(k-1)}\|^2 + \frac{\delta\eta}{L}k^2(F(x^{(k)}) - F(x^*)) = 0.$$

Since the two quantities involved in the above limit are nonnegative, both their limits are equal to 0, which concludes the proof. □

3.2.4. Weak convergence of the iterates to a minimizer. In this section, we prove that the sequence of iterates generated by Algorithm 1 converges to a minimizer. As in [1, Theorem 3], we make use of the improved convergence rates proven in Theorem 3.2 to achieve the result. However, we draw the attention to the fact that the line of proof used in our result greatly simplifies the one in [1, Theorem 3]. Furthermore, Theorem 3.3 improves the result obtained in [11, Theorem 17] for Algorithm 1 with exact proximal evaluations, since here we do not require any further condition on the sequence of operators $\{D_k\}_{k\in\mathbb{N}}$ apart from the hypothesis of summability of the parameters $\{\eta_k\}_{k\in\mathbb{N}}$ in (3.12), whereas in [11, Theorem 17] the operators $\{D_k\}_{k\in\mathbb{N}}$ needed to satisfy an additional condition similar to (3.12) and the parameters $\{\eta_k\}_{k\in\mathbb{N}}$ were required to converge with rate $\mathcal{O}(1/k^p)$, $p > 2$.

Theorem 3.3. *Let Assumptions 1-4 hold and suppose that Assumption 3 holds with $a > 2$. Then*

(i) the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ is bounded and any of its weak limit points is a solution of problem (2.4)*;*
(ii) if, in addition, Assumption 2 holds with $\{D_k\}_{k\in\mathbb{N}} \subseteq \mathcal{D}_\eta^\gamma$, where $\gamma \in \mathbb{R}_{>0}$, $\eta \leq \gamma$, then the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ weakly converges to a solution of problem (2.4)*.*

Proof. (i) The first point follows exactly as in [11, Corollary 15].
(ii) Let x^* be a solution of problem (2.4). We first show that the sequence $\{\|x^{(k)}-x^*\|_{D_{k-1}}\}_{k\in\mathbb{N}}$ converges. If we rewrite the point $z^{(k)}$ from (3.5) as

$$z^{(k)} = x^{(k)} + \frac{k-1}{a}\left(x^{(k)} - x^{(k-1)}\right),$$

and observe that

$$\begin{aligned}\|u^{(k)}\|^2_{D_{k-1}} &= \|z^{(k)} - x^*\|^2_{D_{k-1}}\\ &= \left(\frac{k-1}{a}\right)^2 \|x^{(k)} - x^{(k-1)}\|^2_{D_{k-1}} + 2\left(\frac{k-1}{a}\right)\langle x^{(k)} - x^*, D_{k-1}(x^{(k)} - x^{(k-1)})\rangle\\ &+ \|x^{(k)} - x^*\|^2_{D_{k-1}},\end{aligned} \tag{3.33}$$

then, from Theorem 3.2, the boundedness of $\{x^{(k)}\}_{k\in\mathbb{N}}$ and the fact that $\|x\|^2_{D_{k-1}} \leq \gamma\|x\|^2$ for all $x \in \mathcal{H}$, the first and second terms on the right-hand side of (3.33) converge to 0, so that the sequence $\{\|x^{(k)} - x^*\|_{D_{k-1}}\}_{k\in\mathbb{N}}$ converges if and only if $\{\|u^{(k)}\|_{D_{k-1}}\}_{k\in\mathbb{N}}$ converges.
Starting from inequality (3.15), and recalling the upper bound in (3.2), equation (3.13) which implies that $1/\theta_i \leq i+1$, and the existence of a constant $U \in \mathbb{R}_{>0}$ such that $\|u^{(k)}\|_{D_{k-1}} \leq U$ for all $k \geq 0$, we get

$$\begin{aligned}2\alpha_{k+1}t_k^2 v_{k+1} + \|u^{(k+1)}\|^2_{D_k} &\leq (1+\eta_{k-1})\left(2\alpha_k t_{k-1}^2 v_k + \|u^{(k)}\|^2_{D_{k-1}}\right)\\ &+ 2(k+1)^2\alpha_0\bar{\epsilon}_k + 2U(k+1)\sqrt{2\alpha_0\hat{\epsilon}_k}.\end{aligned} \tag{3.34}$$

Recalling that $\{k^2\bar{\epsilon}_k\}_{k\in\mathbb{N}}$, $\{k\sqrt{\hat{\epsilon}_k}\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ are summable sequences, inequality (3.34) and Lemma 3.2 imply that the sequence $\{2\alpha_k t_{k-1}^2 v_k + \|u^{(k)}\|^2_{D_{k-1}}\}_{k\in\mathbb{N}}$ converges. Furthermore, due to the boundedness of $\{\alpha_k\}_{k\in\mathbb{N}}$, the fact that $t_{k-1}^2 = (k-1+a)^2/a^2$ and Theorem 3.2, also the sequence $\{2\alpha_k t_{k-1}^2 v_k\}_{k\in\mathbb{N}}$ converges. Then it necessarily follows that $\{\|u^{(k)}\|^2_{D_{k-1}}\}_{k\in\mathbb{N}}$ is a convergent sequence and, as previously remarked, this is equivalent to say that $\{\|x^{(k)} - x^*\|_{D_{k-1}}\}_{k\in\mathbb{N}}$ converges.
From this last fact, one can prove that the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ admits a unique weak limit point and, thus, it converges to a solution of problem (2.4), by proceeding exactly as in [11, Theorem 17]. □

We conclude this section by stating an useful corollary which follows from Theorems 3.2 and 3.3.

COROLLARY 3.2. *Let Assumptions 1-3 hold. Suppose that Assumption 2 holds with $\{D_k\}_{k\in\mathbb{N}} \subseteq \mathcal{D}_\eta^\gamma$, where $\gamma \in \mathbb{R}_{>0}$, $\eta \leq \gamma$, Assumption 3 holds with $a > 2$, and one of the following conditions is satisfied:*

- *$\epsilon_k = \mathcal{O}(1/k^p)$, with $p > 4$;*
- *$\hat{\epsilon}_k \equiv 0$ and $\epsilon_k = \bar{\epsilon}_k = \mathcal{O}(1/k^p)$, with $p > 3$.*

Then we have:

- *$F(x^{(k)}) - F(x^*) = o\left(1/k^2\right)$;*
- *$\|x^{(k)} - x^{(k-1)}\| = o\left(1/k\right)$;*
- *$\{x^{(k)}\}_{k\in\mathbb{N}}$ weakly converges to a minimizer of F.*

4. Inexact proximal point computation.

In this section we show in detail how the inexactness criterion (2.7) can be fulfilled in practice for a large class of problems where g has the form

$$g(x) = \sum_{i=1}^{p} \phi_i(M_i x) + \psi(x), \tag{4.1}$$

being $M_i : \mathcal{H} \to \mathcal{Z}_i$ linear bounded operators between Hilbert spaces, $\phi_i : \mathcal{Z}_i \to \mathbb{R}\cup\{\infty\}$, $\psi : \mathcal{H} \to \mathbb{R}\cup\{\infty\}$ proper convex functions. Unlike in [36, 10], we treat the term ψ separately from the functions ϕ_i, instead of setting $\phi_{p+1} = \psi$, $M_{p+1} = \mathcal{I}$; this choice will be better explained afterwards (see Remark 4.1).

4.1. Sufficient conditions to compute an $\epsilon-$approximation.

In the following, we will denote with $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots \times \mathcal{Z}_p$ the Hilbert space equipped with the scalar product $\langle w, v\rangle = \sum_{i=1}^p \langle w_i, v_i\rangle$, where $w = (w_1, ..., w_p) \in \mathcal{Z}$, $v = (v_1, ..., v_p) \in \mathcal{Z}$ are partitioned as $w_i, v_i \in \mathcal{Z}_i$, while $M : \mathcal{H} \to \mathcal{Z}$ denotes the linear bounded operator defined as $Mx = (M_1 x, \cdots, M_p x) \in \mathcal{Z}$.

Dropping for simplicity the iteration index k, the inner subproblem to be solved at the k–th iteration of Algorithm 1 (see Step 3) assumes the form

$$\min_{x\in\mathcal{H}} \mathcal{P}(x) \equiv \sum_{i=1}^{p} \phi_i(M_i x) + \psi(x) + \frac{1}{2\alpha}\|x-\bar{y}\|_D^2, \tag{4.2}$$

where $\bar{y} = y^{(k)} - \alpha_k D_k^{-1}\nabla f(y^{(k)})$, $\alpha \equiv \alpha_k$, $D \equiv D_k$ and $\mathcal{P} \equiv \mathcal{P}_{\alpha,D}$. By exploiting the relation $\phi_i(M_i x) = \max_{w_i\in\mathcal{Z}_i} \langle M_i^* w_i, x\rangle - \phi_i^*(w_i)$, where $\phi_i^* : \mathcal{Z}_i \to \mathbb{R}\cup\{\infty\}$ is the Fenchel conjugate of ϕ_i, we obtain the primal-dual formulation of (4.2)

$$\min_{x\in\mathcal{H}} \max_{w\in\mathcal{Z}} \mathcal{F}(x,w) \equiv \langle M^* w, x\rangle - \sum_{i=1}^{p} \phi_i^*(w_i) + \psi(x) + \frac{1}{2\alpha}\|x-\bar{y}\|_D^2. \tag{4.3}$$

Finally, observing that problem (4.3) is equivalent to

$$\max_{w\in\mathcal{Z}} \min_{x\in\mathcal{H}} \psi(x) + \frac{1}{2\alpha}\|x-(\bar{y}-\alpha D^{-1}M^*w)\|_D^2 - \frac{1}{2\alpha}\|\bar{y}-\alpha D^{-1}M^*w\|_D^2 + \frac{1}{2\alpha}\|\bar{y}\|_D^2 - \sum_{i=1}^{p}\phi_i^*(w_i)$$

and that, by definition (2.5), the unique minimizer of the primal-dual function w.r.t. x is $\mathrm{prox}_{\alpha\psi}^D(\bar{y} - \alpha D^{-1}M^*w)$, we obtain the dual formulation

$$\max_{w\in\mathcal{Z}} \mathcal{Q}(w) \equiv -\sum_{i=1}^{p}\phi_i^*(w_i) + \Phi_{\alpha,D,\bar{y}}(w) \tag{4.4}$$

where $\Phi_{\alpha,D,\bar{y}}(w)$ is defined as

$$\begin{aligned}\Phi_{\alpha,D,\bar{y}}(w) = \psi(\mathrm{prox}_{\alpha\psi}^D(\bar{y}-\alpha D^{-1}M^*w)) + \frac{1}{2\alpha}\|\mathrm{prox}_{\alpha\psi}^D(\bar{y}-\alpha D^{-1}M^*w) - (\bar{y}-\alpha D^{-1}M^*w)\|_D^2 + \\ -\frac{1}{2\alpha}\|\bar{y}-\alpha D^{-1}M^*w\|_D^2 + \frac{1}{2\alpha}\|\bar{y}\|_D^2.\end{aligned}$$

Under the assumption that $0 \in \mathrm{int}(M\mathrm{dom}(\psi) - \mathrm{dom}(\phi_1)\times\cdots\times\mathrm{dom}(\phi_p))$ [35], problems (4.2), (4.3) and (4.4) are equivalent* and, by definition, we have

$$\mathcal{P}(y) \geq \mathcal{F}(y,w) \geq \mathcal{Q}(w) \quad \forall y\in\mathcal{H}, \forall w\in\mathcal{Z}$$

with equalities holding when y and w are solutions of the primal and the dual problem respectively. Since the previous inequalities hold in particular for $y = p_{\alpha,D}(y^{(k)})$, we have

$$\mathcal{P}(x) - \mathcal{P}(p_{\alpha,D}(y^{(k)})) \leq \mathcal{P}(x) - \mathcal{F}(p_{\alpha,D}(y^{(k)}), w) \leq \mathcal{P}(x) - \mathcal{Q}(w).$$

It follows that a sufficient condition for a point x to be an ϵ-approximation fulfilling condition (2.7) is the existence of a dual point w such that $\mathcal{P}(x) - \mathcal{Q}(w) \leq \epsilon$. From a practical point of view, a primal-dual pair (x,w) such that the previous inequality holds can be computed by applying an iterative optimization method to the dual problem, generating a dual sequence $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}} \subseteq \mathcal{Z}$ and a corresponding primal sequence $\{x^{(k,\ell)}\}_{\ell\in\mathbb{N}} \subseteq \mathcal{H}$, and stopping the inner iterations when the previous inequality is satisfied with $w = w^{(k,\ell)}$ and $x = x^{(k,\ell)}$. We formalize the previous remarks in the following proposition.

Proposition 4.1. *Suppose that g has the form (4.1). Let $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ be a sequence in $\mathcal{Z}$ such that*

$$\lim_{\ell\to\infty} \mathcal{Q}(w^{(k,\ell)}) = \bar{\mathcal{Q}}^{(k)} \tag{4.5}$$

*See for example [3, Proposition 15.22] for a more general assumption in general Hilbert spaces, while in finite dimensional spaces a well known assumption guaranteeing the equivalence of the primal, primal-dual and dual problem is $\mathrm{ri}(M\mathrm{dom}(\psi) \cap \mathrm{ri}(\mathrm{dom}(\phi_1)\times\cdots\times\mathrm{dom}(\phi_p))) \neq \emptyset$.

where $\bar{\mathcal{Q}}^{(k)}$ is the maximum of the dual function (4.4)*. Let $\{x^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ satisfying*

$$\lim_{\ell\to\infty}\mathcal{P}(x^{(k,\ell)}) = \mathcal{P}(p_{\alpha,D}(y^{(k)})) = \bar{\mathcal{Q}}^{(k)}. \tag{4.6}$$

Then, for any $\epsilon \geq 0$ and for all sufficiently large ℓ we have

$$\mathcal{P}(x^{(k,\ell)}) - \mathcal{Q}(w^{(k,\ell)}) \leq \epsilon.$$

which implies

$$0 \in \partial_\epsilon \mathcal{P}(x^{(k,\ell)}).$$

The previous proposition gives a general strategy to compute an ϵ-approximation. However, we can further develop this approach to obtain an ϵ-approximation with no errors on gradient computation, by defining the primal sequence in a suitable way, as stated in the next result.

PROPOSITION 4.2. *Suppose that g has the form* (4.1)*, g is continuous on its domain and $dom(g) = dom(\psi)$. Let $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ be any sequence in $\mathcal{Z}$ such that $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ weakly converges to a solution of the dual problem* (4.4) *and* (4.5) *holds. Define the primal sequence as*

$$x^{(k,\ell)} = \mathrm{prox}^{D}_{\alpha\psi}(\bar{y} - \alpha D^{-1}M^* w^{(k,\ell)}), \quad \forall\ \ell \in \mathbb{N}. \tag{4.7}$$

Then, for any $\bar{\epsilon} > 0$ and for any sufficiently large ℓ it holds

$$\mathcal{G}(x^{(k,\ell)}, w^{(k,\ell)}) \equiv \mathcal{P}(x^{(k,\ell)}) - \mathcal{Q}(w^{(k,\ell)}) \leq \bar{\epsilon} \tag{4.8}$$

which implies

$$\frac{1}{\alpha}D(\bar{y} - x^{(k,\ell)}) \in \partial_{\bar{\epsilon}} g(x^{(k,\ell)}).$$

Proof. Since the dual sequence $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ weakly converges to a solution of the dual problem (4.4), and by continuity of the proximity operator $\mathrm{prox}^{D}_{\alpha\psi}$, the primal sequence $\{x^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ weakly converges to $p_{\alpha,D}(y^{(k)})$, which is the solution of (4.2). Thus, the continuity of g implies that

$$\lim_{\ell\to\infty}\mathcal{P}(x^{(k,\ell)}) = \mathcal{P}(p_{\alpha,D}(x^{(k)})) = \bar{\mathcal{Q}}^{(k)}$$

which, together with (4.5) and Proposition 4.1, guarantees that the criterion (4.8) is satisfied for all sufficiently large ℓ. The proof then proceeds by developing the arguments in [36] and in [10]. From the definition of $x^{(k,\ell)}$, we have that

$$\frac{1}{\alpha}D(\bar{y} - x^{(k,\ell)}) - M^* w^{(k,\ell)} \in \partial\psi(x^{(k,\ell)}). \tag{4.9}$$

As a consequence, from the definition of subgradient, we obtain that for any $x \in \mathbb{R}^n$

$$\psi(x) \geq \psi(x^{(k,\ell)}) + \frac{1}{\alpha}\langle D(\bar{y} - x^{(k,\ell)}), x - x^{(k,\ell)}\rangle - \langle M^* w^{(k,\ell)}, x - x^{(k,\ell)}\rangle. \tag{4.10}$$

Moreover, using the definition of the conjugate function we have

$$\begin{aligned}\phi_i^*(w_i^{(k,\ell)}) &= \max_{y_i\in\mathcal{Z}_i}\langle w_i^{(k,\ell)}, y_i\rangle - \phi_i(y_i) \overset{y_i = M_i z}{\geq} \max_{z\in\mathcal{H}}\langle M_i^* w_i^{(k,\ell)}, z\rangle - \phi_i(M_i z)\\ &\geq \langle M_i^* w_i^{(k,\ell)}, x\rangle - \phi_i(M_i x) \quad \forall x \in \mathcal{H}.\end{aligned} \tag{4.11}$$

By computing the gap function $\mathcal{G}(\cdot,\cdot)$ at the pair $(x^{(k,\ell)}, w^{(k,\ell)})$, we obtain

$$\begin{aligned}
\mathcal{G}(x^{(k,\ell)}, w^{(k,\ell)}) &= g(x^{(k,\ell)}) + \sum_{i=1}^{p} \phi^*(w_i^{(k,\ell)}) - \langle M^* w^{(k,\ell)}, x^{(k,\ell)}\rangle - \psi(x^{(k,\ell)}) \\
&\overset{(4.11)}{\geq} g(x^{(k,\ell)}) - \sum_{i=1}^{p} \phi_i(M_i x) + \langle M^* w^{(k,\ell)}, x - x^{(k,\ell)}\rangle - \psi(x^{(k,\ell)}) \\
&\overset{(4.10)}{\geq} g(x^{(k,\ell)}) - \sum_{i=1}^{p} \phi_i(M_i x) - \psi(x) + \frac{1}{\alpha}\langle D(\bar{y} - x^{(k,\ell)}), x - x^{(k,\ell)}\rangle \\
&= g(x^{(k,\ell)}) - g(x) + \frac{1}{\alpha}\langle D(\bar{y} - x^{(k,\ell)}), x - x^{(k,\ell)}\rangle \qquad (4.12)
\end{aligned}$$

where x is any point in $\mathcal{H}$. Rearranging (4.12), from the definition of ϵ-subgradient, we have the result. □

Remark 4.1. *The assumption dom(ψ) = dom(g) implies that $x^{(k,\ell)}$ defined in* (4.7) *belongs to dom(g) for all ℓ. This is crucial, when dom(g) is not the whole space $\mathcal{H}$, to ensure the finiteness of the left-hand side of criterion* (4.8) *and, thus, its well-posedness and practical implementation. In this light, Proposition 4.2 can be seen as a generalization of [10, Proposition 4.1], where the feasibility of the sequence $\{x^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ was not guaranteed.*

4.2. Practical computation of the ϵ-approximation. We now suggest a practical strategy to generate a primal-dual sequence $\{(x^{(k,\ell)}, w^{(k,\ell)})\}_{\ell\in\mathbb{N}}$ satisfying the assumptions of Proposition 4.2.
First, we observe that the objective function of the dual problem (4.4) has the same structure of problem (2.4), since $\Phi_{\alpha,D,\bar{y}}$ is differentiable [26, Prop. 7] and its gradient, given by $\nabla\Phi_{\alpha,D,\bar{y}}(w) = M\text{prox}_{\alpha\psi}^{D}(\bar{y} - \alpha D^{-1}M^*w)$, is Lipschitz–continuous. Using the non-expansivity of the proximity operator, it is immediate to observe that its Lipschitz constant $L_{\Phi_{\alpha,D,\bar{y}}}$ is bounded by $\alpha\|M\|^2\|D^{-1}\|$. Then, to generate the dual sequence $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}}$, we can use the FISTA algorithm as inner solver, setting the fixed inner stepsize as $1/(\alpha\|M\|^2\|D^{-1}\|)$ and the extrapolation parameter as $\beta_k = (k-1)/(k+a)$ with $a > 2$, so that the dual sequence $\{w^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ weakly converges to a maximizer of the dual function and (4.5) holds at an $o(1/k^2)$ rate. The primal sequence $\{x^{(k,\ell)}\}_{\ell\in\mathbb{N}}$ is then computed according to formula (4.7), and both sequences are stopped when condition (4.8) is met.
Regarding the practical realization of the FISTA inner iterates, we observe that the only implicit operations required are the computation of the proximity operators of ϕ_i^* and ψ which, in a wide class of relevant applications as the one considered in Section 5, are available in closed form [14].

5. An application: image deblurring with Poisson noise.

In order to analyze the practical behavior of Algorithm 1, we focus on a specific and relevant image restoration problem, whose variational formulation falls back on the form (2.4). Indeed, in the Bayesian framework, the recovering of an unknown image from a set of noisy data can be obtained by minimizing the sum of a discrepancy function, typically depending on the type of noise affecting the data, plus a regularization term including *a-priori* information and possible physical constraints. In the case of Poisson noise, the discrepancy function measuring the distance from the data $g \in \mathbb{R}^n$ is the generalized Kullback-Leibler (KL) divergence, given by

$$f(x) = KL(Hx + b; z) = \sum_{i=1}^{n} z_i \log \frac{z_i}{(Hx)_i + b} + (Hx)_i + b - z_i \tag{5.1}$$

where the convention $0 \log 0 = 0$ is adopted (see [6] for a detailed survey on image reconstruction from Poisson data). In (5.1), the vector $z \in \mathbb{R}^n$ is the observed image, the matrix $H \in \mathbb{R}^{n\times n}$ represents the blurring operator, while $b \in \mathbb{R}$ is a nonnegative background term. We will assume that H has nonnegative entries and each row and column have at least one positive entry, that is, $He > 0$ and $H^Te > 0$, where $e \in \mathbb{R}^n$ is the vector of all ones. Under these assumptions on H and b, the KL divergence is a nonnegative, convex and coercive function on the nonnegative orthant [18, 7], and explicit expressions of its Hessian and gradient can be computed at the points $x \in Y = \{x \geq 0\} \cap \mathrm{dom} f$; when $b > 0$, Y reduces to the nonnegative orthant and ∇f is Lipschitz continuous on Y [6, 20], although only an upper bound of the Lipschitz constant can be computed. Since the entries of the vector x represent the image pixels, the restriction to Y is required also by the physical motivation that a meaningful solution must be nonnegative. Furthermore, in order to preserve the sharpness of the edges in the reconstructed image, we can consider as regularization term the discrete version of the Total Variation (TV) functional [22]

$$TV(x) = \sum_{i=1}^{n} \|\nabla_i x\|, \tag{5.2}$$

where $\nabla_i \in \mathbb{R}^{2\times n}$ is the discrete gradient operator at the i–th pixel. Thus, the function g in (2.4) has the form

$$g(x) = \rho TV(x) + \iota_Y(x), \tag{5.3}$$

where ι_Y is the indicator function of the nonnegative orthant and ρ is a positive regularization parameter. Obviously $Y = \mathrm{dom} g \subseteq \mathrm{dom} f$. Then, the function (5.3) can be cast in the form (4.1) by setting $\phi_i : \mathbb{R}^2 \to \mathbb{R}$, $\phi_i(x) = \rho\|x\|$, $M_i = \nabla_i$, $i = 1, ..., n$ and $\psi(x) = \iota_Y(x)$ and we can adopt the approach described in Section 4.2 to compute an ϵ-approximation with no errors on the gradient at Step 3 of Algorithm 1 (see Proposition 4.2). Moreover, we have $\phi_i^* = \iota_{B_\rho}$, where $B_\rho \subseteq \mathbb{R}^2$ is the ball centered at the origin with radius ρ; thus the proximity operators of ϕ_i^* and ψ are all Euclidean projections onto B_ρ and Y respectively, which can be computed in closed form.

5.1. Variable metric selection. The sequence of scaling matrices $\{D_k\}_{k\in\mathbb{N}}$ can be selected by exploiting the split gradient idea suggested in [23], which is based on a decomposition of the gradient of the differentiable part into a nonnegative part and a negative one. Here we adopt the same approach used in [10], consisting in applying the splitting to the gradient of the KL function. This approach leads to a diagonal strategy for the variable metric selection where, in particular, the matrices D_k are set as follows

$$D_k = \mathrm{diag}\left(\max\left(\frac{1}{\gamma_k}, \min\left(\gamma_k, \frac{y^{(k)}}{H^T e}\right)\right)\right)^{-1}, \tag{5.4}$$

where the quotient is componentwise and γ_k is a threshold parameter which can be defined in such a way that (3.12) is satisfied. A possible choice is

$$\gamma_k = \sqrt{1 + \frac{t_1}{(k+1)^{t_2}}}, \quad t_1 > 0, \text{ and } t_2 > 1, \tag{5.5}$$

which, as observed in [10, 11], ensures (3.12). Obviously, when $t_1 = 0$, D_k reduces to the identity matrix and the standard Euclidean metric is recovered. In general, it is usual to choose a large value of t_1 to allow more flexibility at the first iterates, while, for large k, D_k converges to the identity matrix at a rate which is controlled by t_2.

The numerical experience in [11] shows that the split gradient strategy underlying the choice (5.4) has good effects on the practical performances of FB methods with extrapolation with exact proximal point computation, i.e., $\epsilon_k = 0$. However, when the proximity operator is computed inexactly, the presence of D_k might affect the performance of the inner solver, since it influences the properties of the dual problem (4.4). In particular, the Lipschitz constant of the gradient of $\Phi_{\alpha,D,\bar{y}}$ in (4.4) for problem (5.1)-(5.3) can be estimated as $8\alpha_k\|D_k^{-1}\|$. Therefore, when FISTA is applied to (4.4) with a fixed steplength equal to $\frac{1}{8\alpha_k\|D_k^{-1}\|}$, large values of the diagonal entries of D_k^{-1} and α_k could lead to short inner steps. Moreover, the range $[1/\gamma_k, \gamma_k]$ is strictly related to the spectrum of the matrix D_k and a large value of γ_k might correspond to an ill conditioned matrix. For example, when some component of $y^{(k)}$ is zero, the condition number of D_k is $\gamma_k \min\{\gamma_k, \max_{i\in\{1,...,n\}} y_i^{(k)}/(H^T e)_i\}$. Thus, on one hand, choosing a large γ_k allows

Table 1
Features of Test problems

problem	reference	n	range	σ_{psf}	b	ρ
phanthom	Shepp-Logan phanthom	256^2	$[0, 1000]$	1.4	10	0.004
cameraman	Matlab	256^2	$[0, 1000]$	1.4	5	0.0091
micro	[37, Figure 8]	128^2	$[1, 69]$	3.2	0.5	0.09

more freedom to select D_k following some performing approach such as the split gradient strategy; on the other side, this might cause the ill conditioning of the inner subproblem, resulting in an increase of the inner iterations number, especially when, as the outer iterations proceed, an high accuracy is required. Thus, the benefits of scaling techniques will be significant in terms of effectiveness only if they are able to counterbalance the complexity required to solve a possibly more demanding inner subproblem. This issue will be numerically investigated in the next section.

5.2. Numerical experiments. We consider a set of three test problems, well known in the literature; the corrupted images have been generated by convolving the original objects with a Gaussian kernel with standard deviation σ_{psf}, adding a constant background and perturbing the obtained blurred images with Poisson noise, simulated through the Matlab `imnoise` function. In Table 1 we report the details of each test problem, including the regularization parameter ρ. In all test problems, we assume reflective boundary conditions, so that the matrix-vector products involving H and H^T are performed via the Discrete Cosine Transform [19]. On the other side, each FISTA inner iteration to solve (4.4) requires the computation of matrix-vector products related to ∇_i and ∇_i^T and projections on B_ρ and onto the nonnegative orthant. For each test problem, we computed an approximate solution x^* of (2.4), which is unique since the objective function is strictly convex, by performing a huge number of iterations by a state-of-the-art method.
The numerical experiments described in this section have been performed in MATLAB (R2015b) on a PC equipped with an Intel(R) Core i7-6500U processor with 2,50 GHz and 16 GB of RAM. The progress of Algorithm 1 towards the solution is evaluated, at each iterate, in terms of the relative difference eF_k between the objective function and the minimum value, and of the relative minimization error ex_k, i.e., the relative distance between the iterates and the solution:

$$eF_k = \frac{F(x^{(k)}) - F(x^*)}{F(x^*)}, \quad ex_k = \frac{\|x^{(k)} - x^*\|}{\|x^*\|}. \tag{5.6}$$

In order to evaluate the effect of the scaling technique on Algorithm 1, in Figures 1-2-3 the sequences $\{eF_k\}_{k\in\mathbb{N}}$ and $\{ex_k\}_{k\in\mathbb{N}}$ are plotted with respect to the iterations and/or computational time, showing the first 20 seconds of run. Since we are interested in evaluating the mutual influence of t_1, t_2 and $\bar{\epsilon}_k$, in this experiment, we fix the rule for selecting $\bar{\epsilon}_k$, while the scaling matrices D_k are chosen as in (5.4)-(5.5) with different values of t_1 and t_2. The other parameters are set as follows.

- *Steplength parameter* α_k. The initial value α_{-1} is set to 10. Typically, it is useful to set this parameter to a 'quite' large value, to prevent too short steps. However, 'too' large values may slow down the performances, since any backtracking step at step 4 of Algorithm 1 requires not only the evaluation of f, but also the computation of an ϵ_k-approximation of $p_{\alpha_k,D_k}(y^{(k)})$. In our experience, we observed that the algorithm self-adjusts this parameter at the very first iterations, possibly reducing it with some backtracking steps (10-12 in the test problems described above), but a reduction of α_k never occurred at the successive iterations.
- *Initialization.* The initial vector $x^{(0)}$ is set equal to z while $w^{(k,0)}$ is set equal to the last inner iterate computed at the iteration $k-1$ (*warmstart* of the inner iterates), except for $k = 0$, where we set $w^{(0,0)} = 0$.
- *The inner tolerance* $\bar{\epsilon}_k$. The tolerance in (4.8) is set as

$$\bar{\epsilon}_k = \min\left(\frac{1}{2}\mathcal{G}(x^{(0)}, 0), \frac{\mathcal{G}(x^{(0)}, 0)}{k^{3.1}}\right).$$

With these settings the assumptions of Corollary 3.2 are satisfied, thus both the weak convergence of the iterates and the convergence of the objective function values at an $o(1/k^2)$ rate are ensured.

- *Other parameters.* The parameter δ at Step 4 is $\frac{1}{1.2}$ while we set $a = 2.1$ in (3.13) for both the outer and inner iterations. Therefore, the convergence of the outer and inner iterates is ensured. Moreover, the observed image z have been rescaled in $[0, 1]$ to eliminate the dependance on the data magnitude.

Figures 1-2-3 show that the variable metric techniques can significantly help to improve the performances of Algorithm 1 with respect to the standard version, also in presence of an inexact computation of the proximity operator. A faster decrease of the objective function is observed with respect to the iteration number in panels (a). Moreover, from panels (b) and (d) we can observe that, for some settings of the parameters t_1 and t_2, the variable metric algorithm outperforms the non scaled one also in terms of computational time. Indeed, allowing a large freedom of choice of the scaling matrix, i.e., choosing a larger value for t_1 and a smaller number for t_2 may speed up the convergence of the outer iterates (see panel (a)), but it might result in a more demanding inner subproblem, requiring a larger number of inner iterations (panel (c)). Conversely, a fast convergence of D_k to the identity might not fully exploit the benefits of the scaling techniques, even if the inner subproblem is easier to solve. With a good setting of t_1 and t_2, the decrease of the outer iterations number and, consequently, of the corresponding number of the matrix-vector products involving H and H^T, is sufficient to counterbalance the increased amount of inner iterations.

In summary, this set of experiments indicates us that a good strategy consists in choosing a ‘large’ value for t_1 and t_2, resulting in more freedom to choose the scaling matrix at the first iterations, when a larger tolerance on the inner subproblem is allowed, before squeezing it quite quickly to the identity matrix in the successive iterations.

As further benchmark, we compare Algorithm 1 with two state-of-the-art methods:

- the preconditioned version of the Chambolle and Pock (CP) primal-dual algorithm [13, 29]. Here we implement the method as described in [17], so that all the required proximity operators can be computed in closed form. According to [29], the preconditioning matrices of the CP method, Σ and T, must satisfy the condition $\|\Sigma^{\frac{1}{2}} A T^{\frac{1}{2}}\|^2 < 1$ where, in our case, the linear operator A is set as the matrix $[H^T, \nabla_1^T, \cdots, \nabla_n^T]^T$; in order to comply with this requirement, we choose Σ and T as the diagonal matrices defined in [29, Equation 10] with $\alpha = 1$;
- the VMILA method [10], a variable metric linesearch based FB scheme with inexact computation of the proximal step, including an implementable stopping criterion for the inner solver; in particular, we used the Matlab code downloadable from `http://www.oasis.unimore.it/site/home/software.html`, combined with the same version of FISTA [4] used for Algorithm 1 as inner solver; in the notation used in [10], we set $\eta = 10^{-6}$, $\alpha_{min} = 10^{-20}$, $\alpha_{max} = \frac{1}{\alpha_{min}}$, $\delta = 0.5$, $\beta = 10^{-4}$, $\gamma = 1$, $t_1 = 10^{10}$ and $t_2 = 2$.

We consider also a non-inertial version of Algorithm 1, named ISTA in the following, in which the inertial step is neglected, while keeping the variable metric, the inexact computation of the proximal step and the same setting for the parameters adopted in the previous experiment. For Algorithm 1 we set $t_1 = 10^{10}$ and $t_2 = 4$ for all test cases.

Figure 4 shows the relative decrease of the objective function values with respect to the number of iterations and the execution time in the first 30 seconds. The comparison with respect to the iterations number is coherent with the theoretical results of the previous sections; furthermore, in spite of the higher computational complexity of each iteration, for a suitable choice of scaling parameters, the performance of Algorithm 1 appears at least comparable and, in most cases, more satisfactory with respect to the other methods.

6. Conclusions and future work.

In this paper we introduced a novel forward–backward method with extrapolation, whose main feature is the combination of the inexact computation of the proximity operator with variable metric techniques. We performed the convergence analysis of the method, proving an $o(1/k^2)$ convergence rate for the objective function values and the convergence of the iterates to a

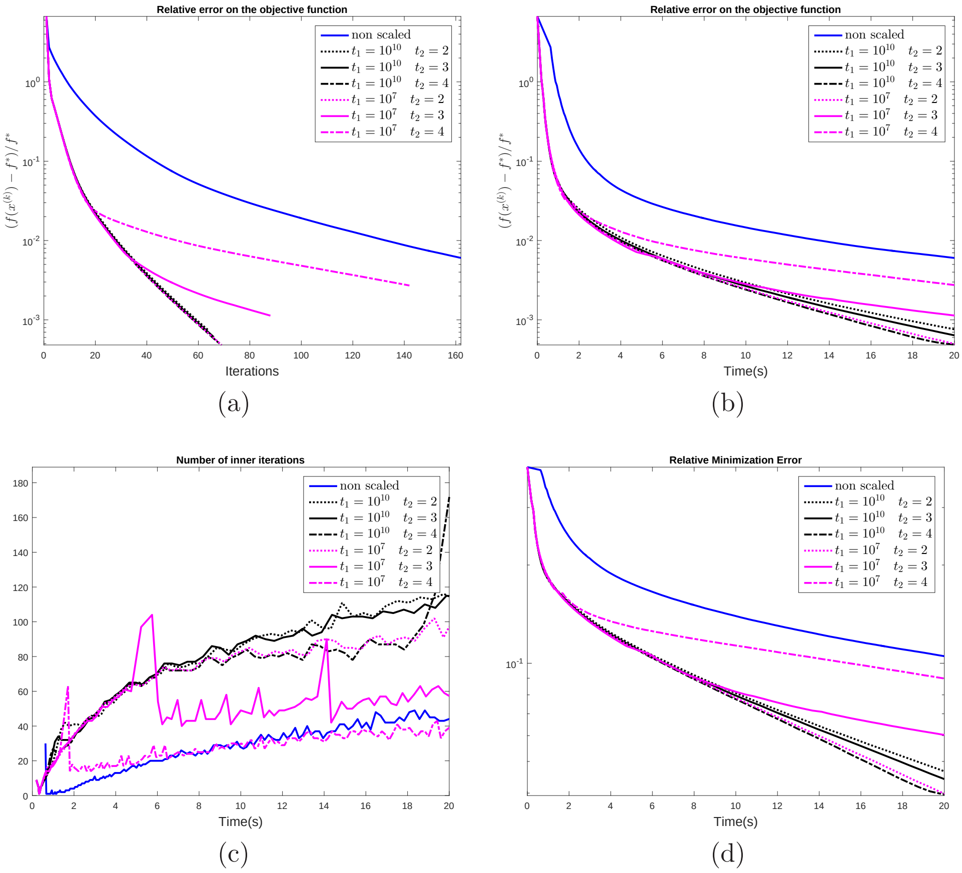


FIG. 1. *Test problem* phantom*: behavior of Algorithm 1 with respect to the scaling techniques. Top row: Relative decrease of the objective function values with respect to the number of iterations (a) and the computational time (b). Bottom row: number of iterations of the inner solver (c) and behavior of the relative minimization error of the iterates with respect to the computational time (d).*

minimizer. These results are obtained by adopting a specific inexactness criterion and under suitable assumptions on the related accuracy and on the variable metric selection. We discussed in detail the implementation of the whole scheme, presenting also the results of a numerical experience on a TV based image restoration problem, where our approach shows to be comparable to other state-of-the-art methods. Further developments of this work can concern the steplength selection rule: indeed, the present algorithm might be inefficient if several reduction of the steplength parameter are required to satisfy the descent lemma, since each of them involves the computation of an approximate proximity operator.

REFERENCES

[1] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$, SIAM J. Optim., 26 (2016), pp. 1824–1834.

[2] H. H. Bauschke, J. Bolte, and M. Teboulle, *A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications*, Math. Oper. Res., 4 (2016), pp. 330–348.

[3] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books on Mathematics, Springer, 2011.

[4] A Beck and M Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Processing, 18 (2009), pp. 2419–34.

[5] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[6] M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini, *Image deblurring with Poisson data: From cells to*

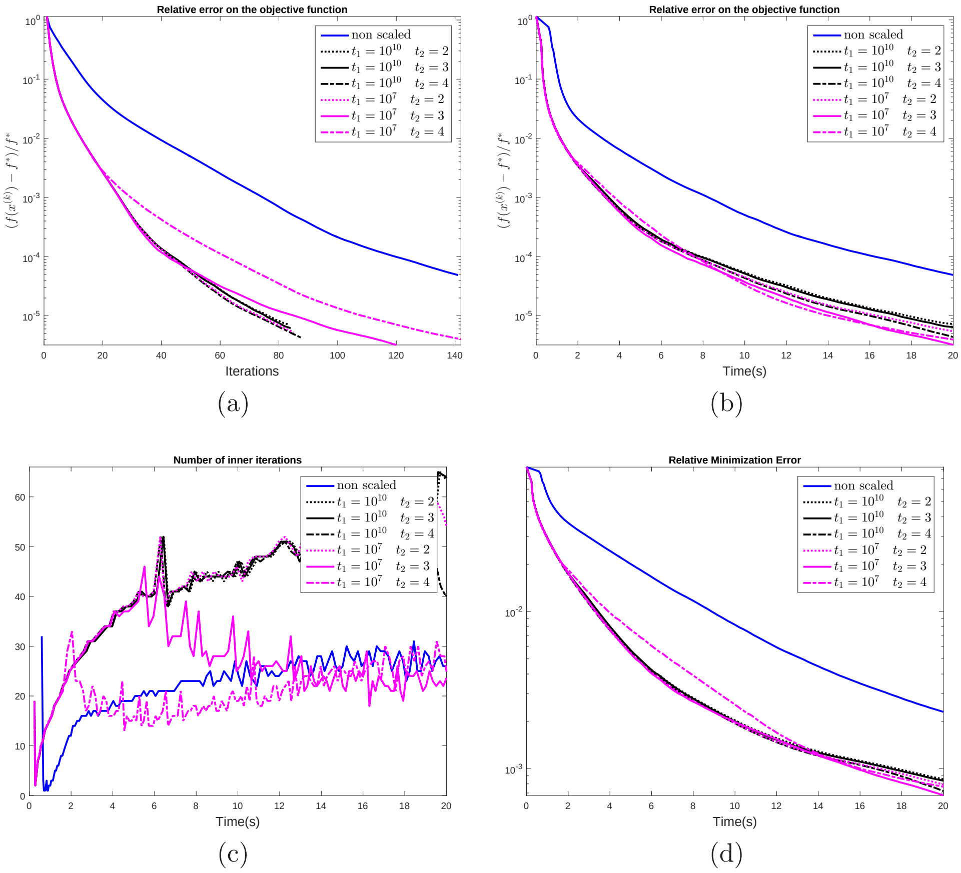


Fig. 2. *Test problem* cameraman*: behavior of Algorithm 1 with respect to the scaling techniques. Top row: Relative decrease of the objective function values with respect to the number of iterations (a) and the computational time (b). Bottom row: number of iterations of the inner solver (c) and behavior of the relative minimization error of the iterates with respect to the computational time (d).*

galaxies, Inverse Probl., 25 (2009), p. 123006.

[7] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, *A discrepancy principle for Poisson data*, Inverse Probl., 26 (2010), p. 105004.

[8] D. Bertsekas, *Convex Optimization Theory*, 5 november 2012 ed., ch. 6 on Convex Optimization Algorithms, pp. 251–489.

[9] ———, *Nonlinear programming*, Athena Scientific, Belmont, 1999.

[10] S. Bonettini, I. Loris, F. Porta, and M. Prato, *Variable metric inexact line–search based methods for nonsmooth optimization*, SIAM J. Optim., 26 (2016), pp. 891–921.

[11] S. Bonettini, F. Porta, and V. Ruggiero, *A variable metric forward-backward method with extrapolation*, SIAM J. Sci. Comput., 38 (2016), pp. A2558–A2584.

[12] A. Chambolle and Ch. Dossal, *On the convergence of the iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm"*, J. Optim. Theory Appl., 166 (2015), pp. 968–982.

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

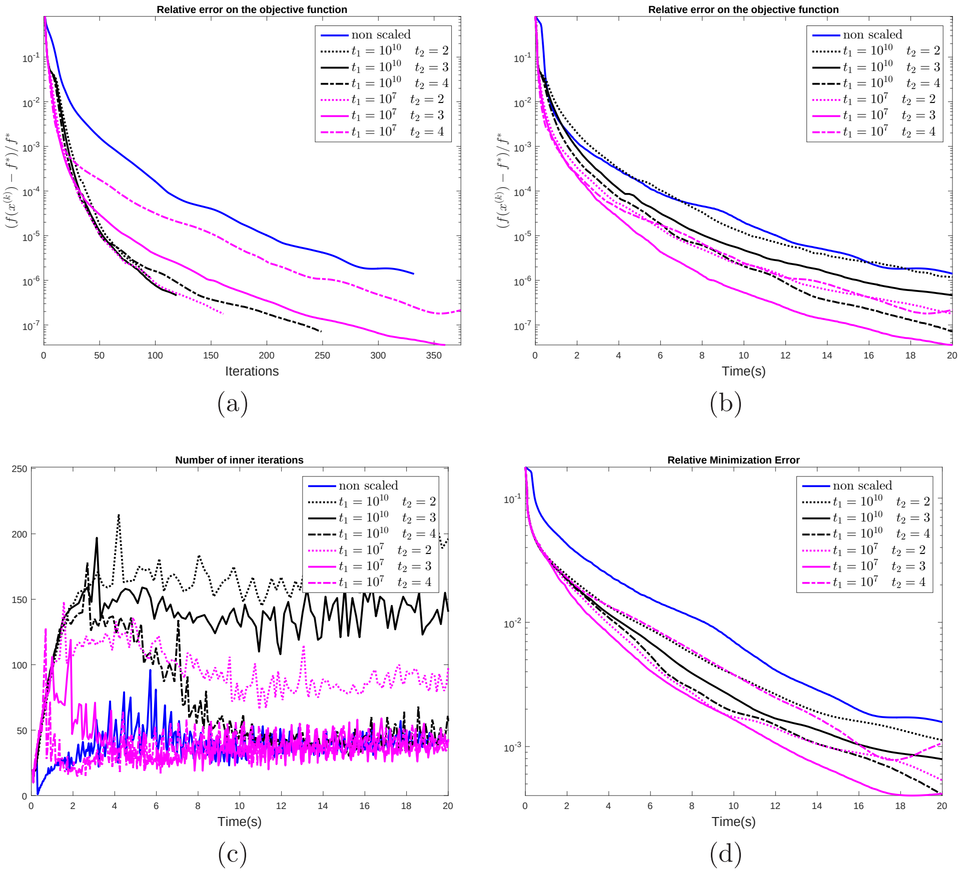
[14] P.L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds., Springer Optimization and Its Applications, Springer, New York NY, 2011, pp. 185–212.

[15] P.L. Combettes and B.C. Vũ, *Variable metric quasi-Féjer monotonicity*, Nonlinear Anal. Theory Methods Appl., 78 (2013), pp. 17–31.

[16] L. Debnath and P. Mikusiński, *Introduction to Hilbert spaces with applications*, Academic Press, Boston, 1990.

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

[18] M. A. T. Figueiredo and J. M. Bioucas-Dias, *Restoration of Poissonian images using alternating direction opti-*



(a) (b)

(c) (d)

FIG. 3. *Test problem* micro*: behavior of Algorithm 1 with respect to the scaling techniques. Top row: Relative decrease of the objective function values with respect to the number of iterations (a) and the computational time (b). Bottom row: number of iterations of the inner solver (c) and behavior of the relative minimization error of the iterates with respect to the computational time (d).*

mization, IEEE Trans. Image Processing, 19 (2010), pp. 3133–3145.

[19] P. C. HANSEN, J. C. NAGY, AND D.P. O'LEARY, *Deblurring Images: Matrices, Spectra and Filtering*, SIAM, Philadelphia, 2005.

[20] Z.T. HARMANY, R.F. MARCIA, AND R.M. WILLETT, *This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms - Theory and practice*, IEEE Trans. Image Processing, 21 (2012), pp. 1084–1096.

[21] J. B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex analysis and minimization algorithms. II*, Springer–Verlag, Berlin, 1993.

[22] RUDIN L, OSHER S, AND FATEMI E, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[23] H. LANTÉRI, M. ROCHE, O. CUEVAS, AND C. AIME, *A general method to devise maximum likelihood signal restoration multiplicative algorithms with non-negativity constraints*, Signal Process., 81 (2001), pp. 945–974.

[24] D. LORENTZ AND T. POCK, *An inertial forward–backward algorithm for monotone inclusion*, J. Math. Imaging Vis., (2014), pp. 311–325.

[25] I. LORIS AND C. VERHOEVEN, *On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty*, Inverse Probl., 27 (2011), p. 125007.

[26] J. J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[27] Y. NESTEROV, *Introductory lectures on convex optimization: a basic course*, Applied optimization, Kluwer Academic Publ., Boston, Dordrecht, London, 2004.

[28] P. OCHS, Y. CHEN, T. BROX, AND T. POCK, *iPiano: Inertial proximal algorithm for non-convex optimization*, SIAM J. Imaging Sci., 7 (2014), pp. 1388–1419.

[29] T. POCK AND A. CHAMBOLLE, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in Computer Vision (ICCV), 2011 IEEE International Conference on, 2011, pp. 1762–1769.

[30] B.T. POLYAK, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys., 4 (1964), pp. 1–17.

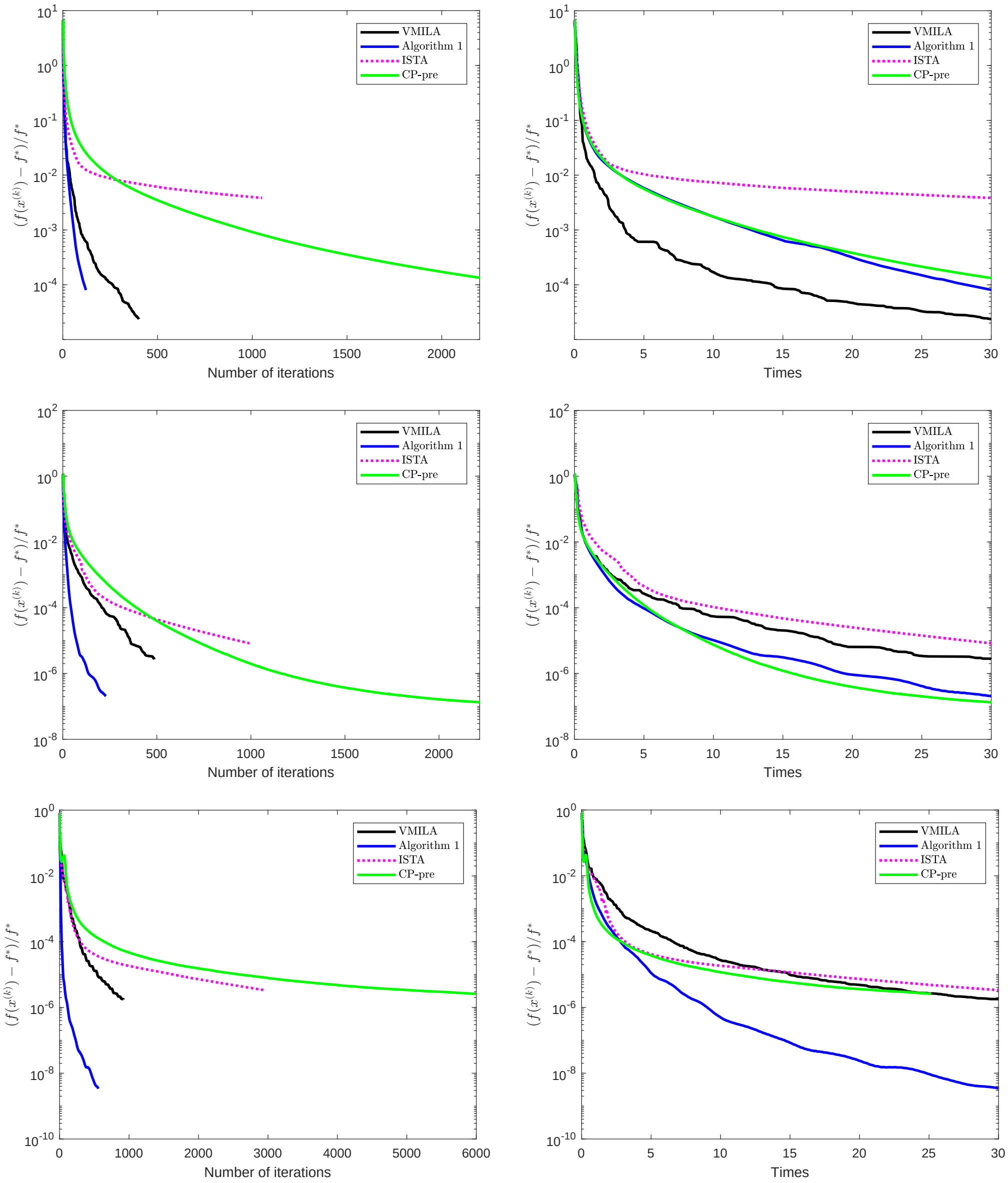


FIG. 4. *Relative decrease of the objective function values with respect to the number of iterations (left column) and the computational time (right column). Top row:* phantom*, middle row:* cameraman*, bottom row:* micro*.*

[31] B. POLYAK, *Introduction to Optimization*, Optimization Software - Inc., Publication Division, N.Y., 1987.

[32] F. PORTA AND I. LORIS, *On some steplength approaches for proximal algorithms*, Appl. Math. Comput., 253 (2015), pp. 345–362.

[33] S. SALZO AND S. VILLA, *Inexact and accelerated proximal point algorithms*, J. Convex Anal., 19 (2012), pp. 1167–1192.

[34] M. SCHMIDT, N. LE ROUX, AND F. BACH, *Convergence rates of inexact proximal-gradient methods for convex optimization*, arXiv:1109.2415v2, (2011).

[35] S. SETZER, *Operator splittings, Bregman methods and frame shrinkage in image processing*, Int. J. Comput. Vis., 92 (2010), pp. 265–280.

[36] S. VILLA, S. SALZO, L. BALDASSARRE, AND A. VERRI, *Accelerated and inexact forward-backward algorithms*, SIAM J. Optim., 23 (2013), pp. 1607–1633.

[37] R. M. WILLETT AND R. D. NOWAK, *Platelets: A multiscale approach for recovering edges and surfaces in photon limited medical imaging*, IEEE Trans. Med. Imaging, 22 (2003), pp. 332–350.

[38] A. ZALINESCU, *Convex analysis in general vector spaces*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.