AN ABSTRACT CONVERGENCE FRAMEWORK WITH APPLICATION TO INERTIAL INEXACT FORWARD–BACKWARD METHODS

SILVIA BONETTINI*, PETER OCHS†, MARCO PRATO†, AND SIMONE REBEGOLDI‡

Abstract. In this paper we introduce a novel abstract descent scheme suited for the minimization of proper and lower semicontinuous functions. The proposed abstract scheme generalizes a set of properties that are crucial for the convergence of several first-order methods designed for nonsmooth nonconvex optimization problems. Such properties guarantee the convergence of the full sequence of iterates to a stationary point, if the objective function satisfies the Kurdyka–Łojasiewicz property. The abstract framework allows for the design of new algorithms. We propose two inertial-type algorithms with (implementable) inexactness criteria for the main iteration update step. The first algorithm, i^{2}Piano, exploits large steps by adjusting a local Lipschitz constant. The second algorithm, iPila, overcomes the main drawback of line-search based methods by enforcing a descent only on a merit function instead of the objective function, which even allows for the escape of local minimizers. Both algorithms are proved to enjoy the full convergence guarantees of the abstract descent scheme. The efficiency of the proposed algorithms is demonstrated on an exemplary image deblurring problem in presence of data corrupted by impulse noise, where we can appreciate the benefits of performing a linesearch along the descent direction inside an inertial scheme.

Key words. forward–backward methods, inertial methods, linesearch, nonconvex optimization

AMS subject classifications. 65K05, 90C30

1. Introduction.

The design of efficient first-order descent methods is vital for tackling composite optimization problems of the form

$$\min_{x\in\mathbb{R}^n} f(x), \quad f(x) = f_0(x) + f_1(x), \tag{1.1}$$

where f_1 is convex and f_0 is continuously differentiable on an open set containing the domain of f_1. Such problems are frequently encountered in image processing and machine learning applications [4, 18, 19], where one of the two terms is usually a data fidelity term and the other one encodes some apriori information on the ground truth [4]. Popular and effective first-order methods aimed at solving (1.1) include forward–backward (FB) methods [25, 24, 31, 12], whose structure consists in the alternation of a gradient step on f_0 followed by a proximal minimization step on f_1, block coordinate methods [9, 14, 23, 26], Douglas-Rachford methods [25, 28] and several others.

In recent years, the convergence of first-order descent methods in nonconvex settings has been carefully addressed by relying on the so-called Kurdyka–Łojasiewicz (KL) inequality [2, 8, 27]. This analytical property is satisfied by a large number of objective functions arising in signal processing and machine learning, such as real analytic or semialgebraic functions (see e.g. [8, 6]), thus making quite natural to consider the KL inequality as a standard blanket assumption whenever the objective function is nonconvex. Combining the KL inequality with some crucial properties of descent methods allows to prove the convergence of the iterates to a stationary point of the objective function, provided that the sequence is bounded. The convergence of descent methods under the KL assumption was first considered in [1], where the authors prove the convergence of linesearch and trust-region methods for real analytic objective functions. The key idea in [1] is to combine the KL inequality with some strong descent conditions holding for the iterates of classical gradient methods. In [2, 9, 26, 10], the authors extend this seminal idea by providing the first *abstract descent schemes* in the KL framework, namely a set of abstract properties ensuring the convergence of a generic iterative scheme to a stationary point if combined with the KL inequality. Such properties include a *sufficient decrease condition* on the function values, a *relative error condition* on the norm of a subgradient at the current iterate, and a *continuity condition* of the iterates with respect to the objective function. In [31, 30], the authors modify the abstract scheme proposed in [2] in order to include an inertial term inside a classical FB splitting scheme and devise the so-called iPiano (inertial Proximal algorithm for nonconvex optimization). Indeed, as

*Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/b, 41125 Modena, Italy (silvia.bonettini@unimore.it, marco.prato@unimore.it).

†Department of Mathematics, University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany (ochs@math.uni-tuebingen.de).

‡Dipartimento di Ingegneria Industriale, Università di Firenze, Via di S. Marta 3, 50139 Firenze, Italy (simone.rebegoldi@unifi.it).

a generalization of the Heavy-Ball method [33, 32], iPiano is not a monotone algorithm, and the sufficient decrease condition cannot be directly imposed on the objective function; however, one can define an appropriate *surrogate function* $\mathcal{F}$ (often called Lyapunov function) in such a way that the decrease condition holds for $\mathcal{F}$ in place of the objective function. Similarly, the authors in [15, 16] define a surrogate function in order to adapt the abstract descent scheme to their proposed method VMILA (Variable Metric Inexact Linesearch Algorithm); yet, unlike in [31], the surrogate function is introduced in order to include an implementable inexactness criterion for the computation of the proximal point, and thus avoid the actual implementation of the relative error condition, which seems rather difficult to impose in practice [15, 29]. Inexactness in FB methods may arise when the proximal operator of the function f_1 cannot be computed in closed form or the exact proximal point is too costly to compute [3, 12, 17, 37].

In this paper, we propose a novel abstract descent scheme for proving the convergence of iterative methods under the KL assumption. The proposed scheme defines two separate surrogate functions, in order to treat separately the possible non–monotonicity of the algorithm and the inexact computation of the iterate, and then imposes a set of abstract properties holding for the surrogate functions evaluated at the iterates. Our approach can be considered as an extension of the abstract scheme in [30], where the inexactness of the iterates is treated by employing the same implementation mechanism used in [16] for VMILA. We show that the iterates generated by our abstract scheme converge to a stationary point of one of the surrogate functions, provided that this function satisfies the KL inequality on its domain. Furthermore, we devise two novel inertial FB algorithms, which are encompassed by the proposed framework. The first one is denominated i^{2}Piano (inertial inexact Proximal algorithm for nonconvex optimization) and can be considered as an inexact version of the iPiano algorithm equipped with a backtracking procedure based on a local version of the Descent Lemma. The second one is denoted iPila (inertial Proximal inexact line–search algorithm), which features an inertial–like step followed by a linesearch procedure along the descent direction of a suitable merit function. The main advantage of iPila resides precisely in its linesearch strategy, which allows to compute the inexact proximal point only once per iteration, unlike the backtracking procedure of i^{2}Piano.

The paper is organized as follows. In Section 2 some basic notions on variational analysis and the definition of the KL property are reported. In Section 3 the proposed abstract scheme is presented and its convergence properties analysed. Section 4 is devoted to the design of the two algorithms i^{2}Piano and iPila and their inclusion in the abstract framework presented in Section 3. Finally, a numerical illustration on an image deblurring problem in presence of impulse noise is reported in Section 5.

2. Preliminaries. In the remainder of the paper, we denote with $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real numbers set and $\mathbb{R}^{n\times n}$ the set of $n \times n$ real-valued matrices, while $\|\cdot\|$ denotes the Euclidean norm. Given a function $\mathcal{F} : \mathbb{R}^n \to \overline{\mathbb{R}}$ and denoting with $\mathrm{dom}(\mathcal{F}) = \{x \in \mathbb{R}^n : \ \mathcal{F}(x) < +\infty\}$ the domain of $\mathcal{F}$, we say that $\mathcal{F}$ is proper if $\mathrm{dom}(\mathcal{F}) \neq \emptyset$ and $\mathcal{F}$ is finite on $\mathrm{dom}(\mathcal{F})$. The distance operator of a point $x \in \mathbb{R}^n$ to a set $\Omega \subset \mathbb{R}^n$ is defined as

$$\mathrm{dist}(x, \Omega) = \inf_{y\in\Omega} \|x - y\|.$$

Observe that, if $\Omega = \Omega_1 \times \Omega_2$, where $\Omega_1 \subset \mathbb{R}^{n_1}$, $\Omega_2 \subset \mathbb{R}^{n_2}$, with $n_1 + n_2 = n$, then for all $x = (x_1, x_2) \in \mathbb{R}^n$, with $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$, we have

$$\mathrm{dist}(x, \Omega) = \sqrt{\mathrm{dist}(x_1, \Omega_1)^2 + \mathrm{dist}(x_2, \Omega_2)^2}. \tag{2.1}$$

This follows by observing that $\|x - z\|^2 = \|x_1 - z_1\|^2 + \|x_2 - z_2\|^2$ for all $z = (z_1, z_2) \in \Omega$.

Definition 2.1. *Given a proper, lower semicontinuous function $\mathcal{F} : \mathbb{R}^n \to \overline{\mathbb{R}}$, the* Fréchet subdifferential *of $\mathcal{F}$ at $\overline{z} \in \mathrm{dom}(\mathcal{F})$ is defined as the set [35, Definition 8.3(a)]*

$$\hat{\partial}\mathcal{F}(\overline{z}) = \left\{ w \in \mathbb{R}^n : \liminf_{u\to\overline{z}, u\neq\overline{z}} \frac{\mathcal{F}(u) - \mathcal{F}(\overline{z}) - (u - \overline{z})^T w}{\|u - \overline{z}\|} \geq 0 \right\}.$$

Furthermore, the limiting subdifferential *of $\mathcal{F}$ at $\overline{z}$ is given by [35, Definition 8.3(b)]*

$$\partial\mathcal{F}(\overline{z}) = \{w \in \mathbb{R}^n : \exists\ z^{(k)} \to \overline{z},\ \mathcal{F}(z^{(k)}) \to \mathcal{F}(\overline{z}),\ w^{(k)} \in \hat{\partial}\mathcal{F}(z^{(k)}) \to w \ \text{as}\ k \to \infty\}.$$

DEFINITION 2.2. *Let $\mathcal{F}:\mathbb{R}^n\to\overline{\mathbb{R}}$ be a proper, lower semicontinuous function. A point $\overline{z}\in\mathbb{R}^n$ is stationary for $\mathcal{F}$ if $0\in\partial\mathcal{F}(\overline{z})$.*

Let us introduce the lazy slope of $\mathcal{F}$ at $\overline{z}$, which is given by [26, p. 877]

$$\|\partial\mathcal{F}(\overline{z})\|_- = \inf_{v\in\partial\mathcal{F}(\overline{z})}\|v\|.$$

It is then easy to prove the following sufficient criterion for establishing if a point $\overline{z}\in\mathbb{R}^n$ is stationary for the function $\mathcal{F}$.

LEMMA 2.3. *[26, Lemma 2.1] Let $\mathcal{F}:\mathbb{R}^n\to\overline{\mathbb{R}}$ be a proper, lower semicontinuous function and $\overline{z}\in\mathbb{R}^n$. If there exists $\{z^{(k)}\}_{k\in\mathbb{N}}\subset\mathbb{R}^n$ such that $z^{(k)}\to\overline{z}$, $\mathcal{F}(z^{(k)})\to\mathcal{F}(\overline{z})$ and $\liminf_{k\to\infty}\|\partial\mathcal{F}(z^{(k)})\|_-=0$, then $0\in\partial\mathcal{F}(\overline{z})$.*

DEFINITION 2.4. *Let $\mathcal{F}:\mathbb{R}^n\longrightarrow\overline{\mathbb{R}}$ be a proper, lower semicontinuous function. The function $\mathcal{F}$ is said to have the KL property at $\overline{z}\in\operatorname{dom}(\partial\mathcal{F})$ if there exist $\upsilon\in(0,+\infty]$, a neighborhood U of $\overline{z}$, a continuous concave function $\phi:[0,\upsilon)\longrightarrow[0,+\infty)$ with $\phi(0)=0$, $\phi\in C^1(0,\upsilon)$, $\phi'(s)>0$ for all $s\in(0,\upsilon)$, such that the following inequality is satisfied*

$$\phi'(\mathcal{F}(z)-\mathcal{F}(\overline{z}))\|\partial\mathcal{F}(z)\|_-\geq 1$$

for all $z\in U\cap\{z\in\mathbb{R}^n:\mathcal{F}(\overline{z})<\mathcal{F}(z)<\mathcal{F}(\overline{z})+\upsilon\}$.
If $\mathcal{F}$ satisfies the KL property at each point of $\operatorname{dom}(\partial\mathcal{F})$, then $\mathcal{F}$ is called a KL function.

The function ϕ in the previous definition is called *desingularization function* and it depends on the point $\bar{z}$. In [9, Lemma 6], the following *uniformized* version of the KL property is introduced, where the KL inequality holds with the same desingularization function for all points in a suitable neighborhood of a compact set where the function is constant.

LEMMA 2.5. *Let $\mathcal{F}:\mathbb{R}^n\to\overline{\mathbb{R}}$ be a proper, lower semicontinuous function and $X\subset\mathbb{R}^n$ a compact set. Suppose that $\mathcal{F}$ satisfies the KL property at each point belonging to X and that $\mathcal{F}$ is constant over X, i.e., $\mathcal{F}(\bar{x})=\bar{\mathcal{F}}\in\mathbb{R}$ for all $\bar{x}\in X$. Then, there exists $\mu,\upsilon>0$ and a function ϕ as in Definition 2.4 such that*

$$\phi'(\mathcal{F}(z)-\bar{\mathcal{F}})\|\partial\mathcal{F}(z)\|_-\geq 1,\quad\forall z\in\bar{B} \tag{2.2}$$

where the set $\bar{B}$ is defined as

$$\bar{B}=\{z\in\mathbb{R}^n:\operatorname{dist}(z,X)<\mu\ \text{and}\ \bar{\mathcal{F}}<\mathcal{F}(z)<\bar{\mathcal{F}}+\upsilon\}. \tag{2.3}$$

3. Abstract algorithm scheme. In the following, we are interested in proving the convergence of an abstract descent algorithm to a stationary point of a proper, lower semicontinuous function $\mathcal{F}$. Such abstract algorithm is defined through a specific set of properties that are shared by several first-order methods designed for nonsmooth nonconvex optimization, including gradient descent methods [1], forward–backward methods [12, 22, 31] and block coordinate methods [2, 14, 26]. Similarly to other abstract descent algorithms in the KL framework, the two main ingredients guaranteeing the convergence of our scheme are the *sufficient decrease condition* and the *relative error condition*, the latter being related to the minimization subproblem that one has to (inexactly) solve at each iteration of a first-order method. However, unlike in previous works in the literature, we require that the relative error condition is satisfied at a point that might be different from the actual iterate generated by the method. As we will see in Section 4, this simple modification allows to circumvent the issue of the actual implementation of the relative error condition, allowing to include inexact forward-backward methods equipped with an implementable inexactness criterion for the solution of the minimization subproblem.

CONDITIONS 3.1 (Abstract algorithm scheme). *Let $\mathcal{F}:\mathbb{R}^n\times\mathbb{R}^m\to\overline{\mathbb{R}}$ be a proper, lower semicontinuous function and $\Phi:\mathbb{R}^n\times\mathbb{R}^q\to\overline{\mathbb{R}}$ a proper, lower semicontinuous, bounded from below function. Consider two sequences $\{x^{(k)}\}_{k\in\mathbb{N}}$, $\{u^{(k)}\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$, a sequence $\{\rho^{(k)}\}_{k\in\mathbb{N}}$ in $\mathbb{R}^m$, a sequence $\{s^{(k)}\}_{k\in\mathbb{N}}$ in $\mathbb{R}^q$ and a sequence of nonnegative real numbers $\{d_k\}_{k\in\mathbb{N}}$ such that the following relations are satisfied.*

[H1] *There exists a sequence of positive real numbers* $\{a_k\}_{k\in\mathbb{N}}$ *such that*

$$\Phi(x^{(k+1)}, s^{(k+1)}) + a_k d_k^2 \leq \Phi(x^{(k)}, s^{(k)}), \quad \forall\ k \geq 0.$$

[H2] *There exists a sequence of nonnegative real numbers* $\{r_k\}_{k\in\mathbb{N}}$ *with* $\lim_{k\to\infty} r_k = 0$ *such that*

$$\Phi(x^{(k+1)}, s^{(k+1)}) \leq \mathcal{F}(u^{(k)}, \rho^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + r_k, \quad \forall\ k \geq 0.$$

[H3] *There exist* $b > 0$*, a sequence of positive real numbers* $\{b_k\}_{k\in\mathbb{N}}$*, a summable sequence of nonnegative real numbers* $\{\zeta_k\}_{k\in\mathbb{N}}$*, a non-empty finite index set* $I \subset \mathbb{Z}$ *and* $\theta_i \geq 0$*,* $i \in I$ *with* $\sum_{i\in I} \theta_i = 1$ *such that, setting* $d_j = 0$ *for* $j \leq 0$*, we have*

$$b_{k+1} \|\partial \mathcal{F}(u^{(k)}, \rho^{(k)})\|_- \leq b \sum_{i\in I} \theta_i d_{k+1-i} + \zeta_{k+1}, \quad \forall\ k \geq 0.$$

[H4] *If* $\{(x^{(k_j)}, \rho^{(k_j)})\}_{j\in\mathbb{N}}$ *is a subsequence of* $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ *converging to some* $(x^*, \rho^*) \in \mathbb{R}^n \times \mathbb{R}^m$*, then we have for* $\{u^{(k_j)}\}_{j\in\mathbb{N}}$*:*

$$\lim_{j\to\infty} \|u^{(k_j)} - x^{(k_j)}\| = 0, \quad \lim_{j\to\infty} \mathcal{F}(u^{(k_j)}, \rho^{(k_j)}) = \mathcal{F}(x^*, \rho^*).$$

[H5] *There exists a positive real number* $p > 0$ *and* $k' \in \mathbb{Z}$ *such that*

$$\|x^{(k+1)} - x^{(k)}\| \leq p d_{k+k'}, \quad \forall\ k \geq 0.$$

[H6] *The sequences* $\{a_k\}_{k\in\mathbb{N}}$*,* $\{b_k\}_{k\in\mathbb{N}}$ *satisfy the following conditions*

$$\sum_{k=0}^{+\infty} b_k = +\infty, \quad \sup_{k\in\mathbb{N}} \frac{1}{b_k a_k} < +\infty, \quad \inf_{k\in\mathbb{N}} a_k > 0.$$

The abstract scheme given in Conditions 3.1 can be seen as a further extension of the one proposed in [30], which is indeed recovered by setting $\Phi = \mathcal{F}$, $u^{(k)} = x^{(k+1)}$ and $\rho^{(k)} \equiv s^{(k)}$. In the following, we discuss in detail conditions *[H1]*-*[H6]* and their relation with the abstract scheme in [30].

- Condition *[H1]* requires the sufficient decrease of a proper, lower semicontinuous function Φ between two successive iterates. The quantity $a_k d_k^2$ measures the amount of the decrease, where d_k is thought as a generalization of the Euclidean norm, whereas the parameter $s^{(k)}$ allows for some flexibility in the asymptotic behaviour of the function Φ. Note that, in earlier works based on the KL property [2, 9, 12, 26], condition *[H1]* is usually presented by setting $d_k = \|x^{(k+1)} - x^{(k)}\|_2$, $\rho^{(k)} \equiv s^{(k)} \equiv 0$ and $\Phi(x, s) = f(x)$, being f the function to minimize. The generalized condition reported here is almost identical to the one introduced in the more recent work [30], with the only difference that here the sufficient decrease is required on a function Φ that may be different from the function $\mathcal{F}$ appearing in *[H3]*.
- Condition *[H3]* is the so-called *relative error condition*, which is related to the (possibly) inexact solution of the minimization subproblem performed at each iteration of a first-order method. In the previous literature [2, 9, 12, 26], such condition is usually employed by setting $u^{(k)} = x^{(k+1)}$, $d_k = \|x^{(k+1)} - x^{(k)}\|_2$, $\rho^{(k)} \equiv 0$, $I = \{1\}$, $\theta_1 = 1$ and $\mathcal{F}(u, \rho) = f(u)$, being f the function to minimize. In [30], a general positive term d_k, a finite index set I, a variable parameter $s^{(k)}$ and a generic surrogate function $\mathcal{F}$ are employed, while keeping $u^{(k)} = x^{(k+1)}$ and $\rho^{(k)} \equiv 0$. Here we also allow the sequence $\{u^{(k)}\}_{k\in\mathbb{N}}$ to be distinct from $\{x^{(k)}\}_{k\in\mathbb{N}}$ and the parameters $\{\rho^{(k)}\}_{k\in\mathbb{N}}$ to vary at each iteration. The reason to do so comes from the fact that condition *[H3]* is hard to enforce algorithmically on $x^{(k+1)}$ when the minimization subproblem is solved inexactly, as noted in [15, 16, 29]. However, if a specific, implementable inexactness criterion is adopted for the solution of the subproblem, then the same condition holds for a surrogate function $\mathcal{F}$ evaluated at a different iterate $(u^{(k)}, \rho^{(k)})$. For instance, this is observed in the convergence analysis of the so-called VMILA algorithm, a variable metric linesearch based forward–backward method studied in [12, 15, 16]. In [12], VMILA is included in the KL framework by setting

$\Phi(x,s) = \mathcal{F}(x,\rho) = f(x)$, $u^{(k)} = x^{(k+1)}$, $d_k = \|x^{(k+1)} - x^{(k)}\|_2$, $\rho^{(k)} \equiv s^{(k)} \equiv 0$ and noting that, in so doing, the relative error condition holds only by exactly computing the proximal operator. In [15], the authors include VMILA in the abstract scheme in a different way, by using $\Phi(x,s) = f(x)$, a surrogate function $\mathcal{F}$ defined upon the concept of forward–backward envelope of f [36], the iterate $u^{(k)}$ as the inexact proximal-gradient point $\tilde{y}^{(k)}$, $d_k = \|x^{(k+1)} - x^{(k)}\|_2$ and $\rho^{(k)}$ as the error parameter due to the computation of $\tilde{y}^{(k)}$. Finally, in [16], VMILA is framed by setting $\Phi(x,s) = f(x)$, $\mathcal{F}(u,\rho) = f(u) + \rho^2/2$, the iterate $u^{(k)}$ as the exact proximal–gradient point $y^{(k)}$ and $d_k = -h^{(k)}(\tilde{y}^{(k)})$, where $h^{(k)}$ is the function to minimize when computing the approximation $\tilde{y}^{(k)}$ of the exact point $y^{(k)}$. In other words, from the analysis in [16], it turns out that we are able to enforce the relative error condition at the exact point $y^{(k)}$, which we do not need to compute explicitly, provided that the approximation $\tilde{y}^{(k)}$ is computed using a specific criterion.

- Condition *[H2]* is crucial in the convergence proof of Theorem 3.3. Indeed it ensures that the sequence $\{\mathcal{F}(u^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$ converges to a limit value $\mathcal{F}^* \in \mathbb{R}$, thus allowing to apply the uniformized KL property at the point $(u^{(k)},\rho^{(k)})$ for all sufficiently large k and, furthermore, it enables the combination of *[H3]* and *[H1]* with the KL inequality. Imposing condition *[H2]* is required only when $\Phi \neq \mathcal{F}$, this is why it does not appear in [30].
- Condition *[H4]* is the analogue of the so-called *continuity condition* in [30]. Here we impose the property for all converging subsequences $(x^{(k_j)},\rho^{(k_j)})$, whereas in [30] it is only required the existence of one such subsequence. This is because, unlike in [30], we need to ensure that the distance between $\{(u^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$ and the limit set of $\{(x^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$ converges to 0 (see Lemma 3.2*(iii)*).
- Condition *[H5]* is also called the *distance condition*. It states the connection between the general term d_k and the Euclidean norm, which is fundamental in order to prove the finite length of the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$. Note that the distance condition given in [30] is slightly more general than *[H5]*; however, that condition alone allows to prove only the finite length of the sequence $\{d_k\}_{k\in\mathbb{N}}$, which in general does not imply the convergence of the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$. In order to obtain the strongest result, condition *[H5]* is then imposed in [30, Theorem 10].
- Condition *[H6]* is the same as the *parameter condition* in [30]. These requirements on the sequences $\{a_k\}_{k\in\mathbb{N}}$, $\{b_k\}_{k\in\mathbb{N}}$ were first introduced in [26] in order to generalize the abstract descent scheme in [2].

In the remainder of this section, we will denote with $\{x^{(k)}\}_{k\in\mathbb{N}}$, $\{u^{(k)}\}_{k\in\mathbb{N}}$, $\{\rho^{(k)}\}_{k\in\mathbb{N}}$, $\{s^{(k)}\}_{k\in\mathbb{N}}$ the sequences complying with Conditions 3.1. Furthermore, let us define the set of all limit points of the sequence $\{(x^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$:

$$\Omega^*(x^{(0)},\rho^{(0)}) = \{(x^*,\rho^*) \in \mathbb{R}^n \times \mathbb{R}^m : \exists\ \{k_j\}_{j\in\mathbb{N}} \subset \mathbb{N} \text{ such that } (x^{(k_j)},\rho^{(k_j)}) \to (x^*,\rho^*)\}.$$

Note that the set $\Omega^*(x^{(0)},\rho^{(0)})$ can be written as

$$\Omega^*(x^{(0)},\rho^{(0)}) = X^*(x^{(0)}) \times R^*(\rho^{(0)})$$

where $X^*(x^{(0)}) = \{x^* \in \mathbb{R}^n : \exists\ \{k_j\}_{j\in\mathbb{N}} \subset \mathbb{N} \text{ such that } x^{(k_j)} \to x^*\} \subset \mathbb{R}^n$, $R^*(\rho^{(0)}) = \{\rho^* \in \mathbb{R}^m : \exists\ \{k_j\}_{j\in\mathbb{N}} \subset \mathbb{N} \text{ such that } \rho^{(k_j)} \to \rho^*\} \subset \mathbb{R}^m$.

LEMMA 3.2. *Let Conditions 3.1 be satisfied. Suppose that $\{(x^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$ is a bounded sequence. Then the following facts hold true.*

(i) $\Omega^*(x^{(0)},\rho^{(0)})$ is nonempty and compact.

(ii) There exists $\mathcal{F}^* \in \mathbb{R}$ such that $\lim_{k\to\infty} \Phi(x^{(k)},s^{(k)}) = \lim_{k\to\infty} \mathcal{F}(u^{(k)},\rho^{(k)}) = \mathcal{F}^*$.

(iii) We have

$$\lim_{k\to\infty} \operatorname{dist}((x^{(k)},\rho^{(k)}),\Omega^*(x^{(0)},\rho^{(0)})) = \lim_{k\to\infty} \operatorname{dist}((u^{(k)},\rho^{(k)}),\Omega^*(x^{(0)},\rho^{(0)})) = 0.$$

(iv) We have $\mathcal{F}(x^*,\rho^*) = \mathcal{F}^*$, $\forall\ (x^*,\rho^*) \in \Omega^*(x^{(0)},\rho^{(0)})$.

Proof. (i) Since the sequence $\{(x^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$ is bounded, it admits at least a limit point and, hence, $\Omega^*(x^{(0)},\rho^{(0)})$ is nonempty. Compactness can be proved by observing that $\Omega^*(x^{(0)},\rho^{(0)})$ is

a countable intersection of compact sets (see [9, Lemma 5]).
(ii) From *[H1]* we have that the sequence $\{\Phi(x^{(k)}, s^{(k)})\}_{k\in\mathbb{N}}$ is nonincreasing and, since Φ is bounded from below, there exists $\mathcal{F}^* \in \mathbb{R}$ such that

$$\lim_{k\to\infty} \Phi(x^{(k)}, s^{(k)}) = \mathcal{F}^*.$$

The previous relation combined with *[H2]* proves Part *(ii)*.
(iii) Since $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ is bounded and by definition of $\Omega^*(x^{(0)}, \rho^{(0)})$, we have

$$\lim_{k\to\infty} \operatorname{dist}((x^{(k)}, \rho^{(k)}), \Omega^*(x^{(0)}, \rho^{(0)})) = 0.$$

Observing that $\Omega^*(x^{(0)}, \rho^{(0)}) = X^*(x^{(0)}) \times R^*(\rho^{(0)})$ and recalling (2.1), the previous limit implies

$$\lim_{k\to\infty} \operatorname{dist}(x^{(k)}, X^*(x^{(0)})) = 0, \qquad \lim_{k\to\infty} \operatorname{dist}(\rho^{(k)}, R^*(\rho^{(0)})) = 0.$$

Combining the boundedness of $\{x^{(k)}\}_{k\in\mathbb{N}}$ with the definition of $X^*(x^{(0)})$ and property *[H4]*, we obtain $\lim_{k\to\infty} \operatorname{dist}(u^{(k)}, X^*(x^{(0)})) = 0$, which together with the second limit above yields

$$\lim_{k\to\infty} \operatorname{dist}((u^{(k)}, \rho^{(k)}), \Omega^*(x^{(0)}, \rho^{(0)})) = 0.$$

(iv) This point follows directly from part *(ii)* and *[H4]*. □

THEOREM 3.3. *Let Conditions 3.1 be satisfied, and suppose that $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ is a bounded sequence and that $\mathcal{F}$ is a KL function. Then the following statements are true.*

(i) The sequence $\{d_k\}_{k\in\mathbb{N}}$ is summable, i.e., it satisfies

$$\sum_{k=0}^{+\infty} d_k < +\infty.$$

(ii) The sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ has finite length, i.e., it satisfies

$$\sum_{k=0}^{+\infty} \|x^{(k+1)} - x^{(k)}\| < +\infty$$

and thus $\{x^{(k)}\}_{k\in\mathbb{N}}$ is a convergent sequence.

(iii) If also $\{\rho^{(k)}\}_{k\in\mathbb{N}}$ converges, then the sequence $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ converges to a stationary point for $\mathcal{F}$.

Proof. (i) By Lemma 3.2*(i)-(iv)*, the function $\mathcal{F}$ is constant over the compact set $\Omega^*(x^{(0)}, \rho^{(0)})$, therefore we can apply Lemma 2.5. Let $\upsilon, \mu, \phi, \bar{B}$ be as in Lemma 2.5. Thanks to Lemma 3.2*(ii)-(iii)* and *[H2]*, there exists a positive integer k_0 such that

$$\Phi(x^{(k)}, s^{(k)}) + r_k < \mathcal{F}^* + \upsilon, \quad \operatorname{dist}((u^{(k)}, \rho^{(k)}), \Omega^*(x^{(0)}, \rho^{(0)})) < \mu \tag{3.1}$$

for all $k \geq k_0$. Without loss of generality, up to a translation of the iteration index, we can assume $k_0 = 0$.
Let us now set $c = \sup_k 1/(a_k b_k)$, where $c < +\infty$ due to *[H6]*, $\zeta'_k = \zeta_k / b$ and

$$\phi_k = \frac{b}{c}(\phi(\Phi(x^{(k)}, s^{(k)}) - \mathcal{F}^*) - \phi(\Phi(x^{(k+1)}, s^{(k+1)}) - \mathcal{F}^*))$$

and prove that

$$2d_k \leq \phi_k + \sum_{i\in I} \theta_i d_{k-i} + \zeta'_k, \quad \forall\ k \geq 1. \tag{3.2}$$

We first observe that the definition of ϕ_k is well posed, since $\mathcal{F}^*$ is the limit of the nonincreasing sequence $\{\Phi(x^{(k)}, s^{(k)})\}_{k\in\mathbb{N}}$ (see Lemma 3.2) and, hence, we have $\Phi(x^{(k)}, s^{(k)}) \geq \mathcal{F}^*$, $\forall\ k \in \mathbb{N}$. Moreover, the monotonicity of both function ϕ and sequence $\{\Phi(x^{(k)}, s^{(k)})\}_{k\in\mathbb{N}}$ implies that $\phi_k \geq 0$, $\forall\ k \in \mathbb{N}$.
Let us now consider the two cases $d_k = 0$ and $d_k > 0$ separately.
If $d_k = 0$, inequality (3.2) holds trivially. Otherwise, for any iteration index $k \geq 1$ such that $d_k > 0$, taking into account *[H1]* and *[H2]*, we can write

$$\mathcal{F}^* \leq \Phi(x^{(k+1)}, s^{(k+1)}) < \Phi(x^{(k)}, s^{(k)}) \leq \mathcal{F}(u^{(k-1)}, \rho^{(k-1)}).$$

On the other hand, the rightmost inequality in *[H2]* gives

$$\mathcal{F}(u^{(k-1)}, \rho^{(k-1)}) \leq \Phi(x^{(k-1)}, s^{(k-1)}) + r_{k-1},$$

which, in view of (3.1), implies that $(u^{(k-1)}, \rho^{(k-1)}) \in \bar{B}$. Then, we can write the KL inequality related to the function $\mathcal{F}$ at $(u^{(k-1)}, \rho^{(k-1)})$:

$$\phi'(\mathcal{F}(u^{(k-1)}, \rho^{(k-1)}) - \mathcal{F}^*) \geq \frac{1}{\|\partial\mathcal{F}(u^{(k-1)}, \rho^{(k-1)})\|_-}.$$

Furthermore, combining the previous inequality with *[H3]* yields

$$\phi'(\mathcal{F}(u^{(k-1)}, \rho^{(k-1)}) - \mathcal{F}^*) \geq \frac{1}{\frac{b}{b_k}\sum_{i\in I}\theta_i d_{k-i} + \frac{1}{b_k}\zeta_k}.$$

Since ϕ is concave, ϕ' is nonincreasing. Therefore, *[H2]* implies

$$\phi'(\Phi(x^{(k)}, s^{(k)}) - \mathcal{F}^*) \geq \phi'(\mathcal{F}(u^{(k-1)}, \rho^{(k-1)}) - \mathcal{F}^*).$$

Exploiting again the concavity of ϕ, we have

$$\phi(\Phi(x^{(k)}, s^{(k)})-\mathcal{F}^*)-\phi(\Phi(x^{(k+1)}, s^{(k+1)})-\mathcal{F}^*) \geq \phi'(\Phi(x^{(k)}, s^{(k)})-\mathcal{F}^*)(\Phi(x^{(k)}, s^{(k)})-\Phi(x^{(k+1)}, s^{(k+1)})).$$

Combining the last three relations with *[H1]* leads to

$$\begin{aligned}\phi(\Phi(x^{(k)}, s^{(k)}) - \mathcal{F}^*) - \phi(\Phi(x^{(k+1)}, s^{(k+1)}) - \mathcal{F}^*) &\geq \frac{\Phi(x^{(k)}, s^{(k)}) - \Phi(x^{(k+1)}, s^{(k+1)})}{\frac{b}{b_k}\sum_{i\in I}\theta_i d_{k-i} + \frac{1}{b_k}\zeta_k} \\ &\geq \frac{a_k d_k^2}{\frac{b}{b_k}\sum_{i\in I}\theta_i d_{k-i} + \frac{1}{b_k}\zeta_k}.\end{aligned}$$

Recalling the definition of ϕ_k and ζ_k', the above inequality implies the following one

$$d_k^2 \leq \phi_k\left(\sum_{i\in I}\theta_i d_{k-i} + \zeta_k'\right).$$

Taking the square root of both sides and using the inequality $2\sqrt{uv} \leq u+v$ on the right-hand-side, we obtain (3.2).
Summing (3.2) from 1 to k leads to

$$2\sum_{j=1}^{k} d_j \leq \sum_{j=1}^{k}\phi_j + \sum_{j=1}^{k}\sum_{i\in I}\theta_i d_{j-i} + \sum_{j=1}^{k}\zeta_j'. \tag{3.3}$$

We now observe that

$$\begin{aligned}\sum_{j=1}^{k}\phi_j &= \frac{b}{c}(\phi(\Phi(x^{(1)}, s^{(1)}) - \mathcal{F}^*) - \phi(\Phi(x^{(k+1)}, s^{(k+1)}) - \mathcal{F}^*)) \\ &\leq \frac{b}{c}\phi(\Phi(x^{(1)}, s^{(1)}) - \mathcal{F}^*),\end{aligned}$$

where the rightmost inequality follows from the positive sign of ϕ. Furthermore, the second sum in the right-hand side of (3.3) can be rewritten as below

$$\begin{aligned}\sum_{j=1}^{k}\sum_{i\in I}\theta_i d_{j-i} &= \sum_{i\in I}\sum_{j=1}^{k}\theta_i d_{j-i} = \sum_{i\in I}\sum_{r=1-i}^{k-i}\theta_i d_r \\ &\leq \sum_{i\in I}\sum_{r=1-i}^{0}\theta_i d_r + \left(\sum_{i\in I}\theta_i\right)\sum_{r=1}^{k} d_r + \sum_{i\in I}\sum_{r=k+1}^{k-i}\theta_i d_r \\ &= \sum_{i\in I}\sum_{r=1-i}^{0}\theta_i d_r + \sum_{r=1}^{k} d_r + \sum_{i\in I}\sum_{r=k+1}^{k-i}\theta_i d_r\end{aligned}$$

where we have used the change of variable $r = j-1$ and the property $\sum_{i\in I}\theta_i = 1$. Note that the sums appearing in the previous relation are assumed to be zero whenever the start index of the summation is larger than the termination index. Therefore we can write

$$2\sum_{j=1}^{k} d_j \leq \sum_{i\in I}\sum_{r=1-i}^{0}\theta_i d_r + \sum_{r=1}^{k} d_r + \sum_{i\in I}\sum_{r=k+1}^{k-i}\theta_i d_r + \frac{b}{c}\phi(\Phi(x^{(1)}, s^{(1)}) - \mathcal{F}^*) + \sum_{j=1}^{k}\zeta'_j$$

which clearly implies

$$\sum_{j=1}^{k} d_j \leq \sum_{i\in I}\sum_{r=1-i}^{0}\theta_i d_r + \sum_{i\in I}\sum_{r=k+1}^{k-i}\theta_i d_r + \frac{b}{c}\phi(\Phi(x^{(1)}, s^{(1)}) - \mathcal{F}^*) + \sum_{j=1}^{k}\zeta'_j. \tag{3.4}$$

At this point, observe that the first two sums in the right-hand side of (3.4) are finite linear combinations of the terms $\{d_r\}_{r\in\mathbb{N}}$. Conditions *[H1]* and *[H6]* ensure that $d_k \to 0$, hence those sums are converging to 0 for $k \to \infty$. Noting also that $\{\zeta'_k\}_{k\in\mathbb{N}}$ is summable and taking the limit of (3.4) for $k \to \infty$, we obtain

$$\sum_{k=0}^{\infty} d_k < +\infty. \tag{3.5}$$

(ii) Combining (3.5) with *[H5]*, we also obtain

$$\sum_{k=0}^{\infty}\|x^{(k+1)} - x^{(k)}\| \leq p\sum_{k=0}^{\infty} d_{k+k'} < +\infty.$$

which implies that the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to a point $x^* \in \mathbb{R}^n$.
(iii) Let $(x^*, \rho^*) \in \mathbb{R}^n \times \mathbb{R}^m$ be the unique limit point of the sequence $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$, namely $(x^{(k)}, \rho^{(k)}) \to (x^*, \rho^*)$. By using *[H4]*, it follows that $(u^{(k)}, \rho^{(k)}) \to (x^*, \rho^*)$ and $\mathcal{F}(u^{(k)}, \rho^{(k)}) \to \mathcal{F}(x^*, \rho^*)$. Furthermore, summing *[H3]* for $k = 0, \dots, K$ yields

$$\sum_{k=0}^{K} b_{k+1}\|\partial\mathcal{F}(u^{(k)}, \rho^{(k)})\|_- \leq b\sum_{k=0}^{K}\sum_{i\in I}\theta_i d_{k+1-i} + \sum_{k=0}^{K}\zeta_k.$$

Taking the limit for $K \to \infty$, using (3.5) and recalling that $\{\zeta_k\}_{k\in\mathbb{N}}$ is summable, we obtain

$$\sum_{k=0}^{\infty} b_{k+1}\|\partial\mathcal{F}(u^{(k)}, \rho^{(k)})\|_- < +\infty.$$

Since *[H6]* requires $\sum_k b_k = +\infty$, the previous relation implies that

$$\liminf_{k\to\infty}\|\partial\mathcal{F}(u^{(k)}, \rho^{(k)})\|_- = 0.$$

In conclusion, the sequence $\{(u^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ satisfies all the hypotheses of Lemma 2.3, which means that $0 \in \partial\mathcal{F}(x^*, \rho^*)$.

□

4. Applications of the abstract scheme. In this section, we show how we can devise some brand new forward–backward–type algorithms satisfying Conditions 3.1 and, hence, guarantee their convergence to a stationary point in virtue of Theorem 3.3. In particular, from now on, we address the problem

$$\min_{x\in\mathbb{R}^n} f(x), \quad f(x) = f_0(x) + f_1(x), \tag{4.1}$$

where we assume that f_0, f_1 are as follows:

[A1] $f_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, lower semicontinuous, convex function;

[A2] $f_0 : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable on an open set $\Omega_0 \supset \overline{\mathrm{dom}(f_1)}$.

[A3] f_0 has $L-$Lipschitz continuous gradient on $\mathrm{dom}(f_1)$, i.e.,

$$\|\nabla f_0(x) - \nabla f_0(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathrm{dom}(f_1),$$

for some $L > 0$.

[A4] f is bounded from below.

Under the above assumptions, for any $z \in \mathrm{dom}(f_1)$, the following subdifferential calculus rules hold [35, Proposition 8.12, Exercise 8.8(c)]

$$\begin{aligned} \partial f_1(z) &= \{w \in \mathbb{R}^n : f_1(y) \geq f_1(z) + \langle w, y - z\rangle, \ \forall y \in \mathbb{R}^n\} \\ \partial f(z) &= \{\nabla f_0(z)\} + \partial f_1(z). \end{aligned} \tag{4.2}$$

One of the most popular algorithms for solving problem (4.1) is the inertial, or heavy ball, proximal-gradient method iPiano [31, 30], which is defined by the iteration

$$x^{(k+1)} = \mathrm{prox}_{\alpha_k f_1}(x^{(k)} - \alpha_k \nabla f_0(x^{(k)}) + \beta_k(x^{(k)} - x^{(k-1)})), \tag{4.3}$$

where α_k, β_k are suitably chosen parameters. By definition of the proximity operator, the above updating rule consists in defining the new point as the unique solution of the minimization problem

$$\min_{y\in\mathbb{R}^n} f_1(y) - f_1(x^{(k)}) + \langle \nabla f_0(x^{(k)}) - \frac{\beta_k}{\alpha_k}(x^{(k)} - x^{(k-1)}), y - x^{(k)}\rangle + \frac{1}{2\alpha_k}\|y - x^{(k)}\|^2. \tag{4.4}$$

We will refer to the minimizer of this problem as the *inertial proximal gradient point*. If $\beta_k = 0$, we recover the standard proximal gradient point, otherwise the inertial step $x^{(k)} - x^{(k-1)}$ is included in the argument of the proximal gradient operator, with the aim of improving the convergence behaviour of the overall method.

In the following, we address the key challenge of designing algorithms that inexactly compute the inertial proximal gradient point with implementable conditions that still preserve the convergence guarantees of Theorem 3.3. More precisely, we propose two new inexact inertial–type algorithms, where the second one also features a linesearch procedure along a descent direction of a suitable merit function. The convergence analysis of both algorithms can be performed in the abstract framework provided by Conditions 3.1. We stress that, due to the inexactness in the computation of the inertial proximal gradient point, our algorithms cannot be cast in the abstract frameworks proposed in previous works.

We start our presentation by defining the inexactness criterion for the inertial proximal gradient point, which is a generalization of the one proposed in [11, 12, 15].

4.1. Inexact inertial proximal gradient point. Given two positive parameters α, β, consider the function $h : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined as follows:

$$h(y; x, s) = f_1(y) - f_1(x) + \langle \nabla f_0(x) - \frac{\beta}{\alpha}(x - s), y - x\rangle + \frac{1}{2\alpha}\|y - x\|^2. \tag{4.5}$$

Clearly, the inertial proximal gradient point (4.3) is the minimizer of the above function with respect to the first argument, with $\alpha = \alpha_k$, $\beta = \beta_k$, $x = x^{(k)}$, $s = x^{(k-1)}$.

Given (x, s), we denote by $\hat{y}$ the (exact) minimizer of the function in (4.5)

$$\hat{y} = \operatorname*{argmin}_{y\in\mathbb{R}^n} h(y; x, s) = \mathrm{prox}_{\alpha f_1}(x - \alpha\nabla f_0(x) + \beta(x - s)). \tag{4.6}$$

The point $\hat{y}$ is the unique point satisfying the optimality condition

$$0 \in \partial h(\hat{y}; x, s) \Leftrightarrow -\frac{1}{\alpha}(\hat{y} - x + \alpha \nabla f_0(x) - \beta(x - s)) \in \partial f_1(\hat{y}). \tag{4.7}$$

Borrowing the ideas in [11, 16], we define an approximation of $\hat{y}$ as any point $\tilde{y} \in \mathrm{dom}(f_1)$ such that

$$h(\tilde{y}; x, s) - h(\hat{y}; x, s) \leq -\frac{\tau}{2} h(\tilde{y}; x, s). \tag{4.8}$$

The above condition is equivalent to the following one:

$$h(\tilde{y}; x, s) \leq \left(\frac{2}{2+\tau}\right) h(\hat{y}; x, s) \leq 0, \tag{4.9}$$

where the rightmost inequality is a consequence of the fact that $\hat{y}$ is a minimizer of $h(\cdot; x, s)$ and $h(x; x, s) = 0$. Therefore, we have $h(\tilde{y}; x, s) \leq 0$ and condition (4.8) can be rewritten in equivalent way as

$$0 \in \partial_\epsilon h(\tilde{y}; x, s), \quad \text{with } \epsilon = -\frac{\tau}{2} h(\tilde{y}; x, s), \tag{4.10}$$

where $\partial_\epsilon h(\tilde{y}; x, s) = \{w \in \mathbb{R}^n : h(y; x, s) \geq h(\tilde{y}; x, s) + \langle w, y - \tilde{y}\rangle - \epsilon,\ \forall y \in \mathbb{R}^n\}$ is the ϵ-subdifferential in the first argument of the convex function $h(\cdot; x, s)$ at point $\tilde{y}$ [38, p. 82]. Therefore, the point $\tilde{y}$ is defined upon a relaxation of the optimality condition (4.7), where the subdifferential of $h(\cdot; x, s)$ is replaced by the ϵ-subdifferential and the accuracy parameter ϵ is chosen in a specific way, which is crucial for preserving the theoretical convergence properties.
Even if the inclusion (4.10) is implicit, a point $\tilde{y} \in \mathrm{dom}(f_1)$ satisfying (4.10) can be actually computed in practice with a well defined, explicit primal–dual procedure, as explained in [11, 17]: an iterative optimization method is applied to the dual of the function $h(\cdot; x, s)$, until the difference between the primal and the dual function is below the tolerance ϵ. A more detailed discussion about the inexactness criterion and on its practical realization can be found also in [16, Section 3.1].

4.1.1. Preliminary results. We collect new several basic results to incorporate inexactness into the design and analysis of our proposed algorithms. The following lemma is a consequence of the strong convexity of the function $h(\cdot; x, s)$, and will be often employed in the following.

LEMMA 4.1. *Suppose that Assumptions [A1]–[A2] hold true. For a given pair* $(x, s) \in \mathrm{dom}(f_1) \times \mathbb{R}^n$*, let* $\hat{y}, \tilde{y}$ *be defined as in* (4.6),(4.8)*. Then, the following inequalities hold.*

$$\frac{1}{2\alpha}\|\hat{y} - x\|^2 \leq \left(1 + \frac{\tau}{2}\right)(-h(\tilde{y}; x, s)) \tag{4.11}$$

$$\frac{1}{2\alpha}\|\tilde{y} - \hat{y}\|^2 \leq \frac{\tau}{2}(-h(\tilde{y}; x, s)) \tag{4.12}$$

$$\frac{\theta}{2\alpha}\|\tilde{y} - x\|^2 \leq (-h(\tilde{y}; x, s)), \quad \textit{with } \theta = 1/\left(\sqrt{1 + \frac{\tau}{2}} + \sqrt{\frac{\tau}{2}}\right)^2 \leq 1. \tag{4.13}$$

Proof. Inequalities (4.11)–(4.12) follow by combining the strong convexity of the function $h(\cdot; x, s)$ and condition (4.8) as in [16, Lemma 2]. As for (4.13), we have

$$\begin{aligned}
\frac{1}{2\alpha}\|\tilde{y} - x\|^2 &= \frac{1}{2\alpha}\|\tilde{y} - \hat{y} + \hat{y} - x\|^2 = \frac{1}{2\alpha}\|\tilde{y} - \hat{y}\|^2 + \frac{1}{2\alpha}\|\hat{y} - x\|^2 + \frac{1}{\alpha}\langle \tilde{y} - \hat{y}, \hat{y} - x\rangle \\
&\leq \frac{1}{2\alpha}\|\tilde{y} - \hat{y}\|^2 + \frac{1}{2\alpha}\|\hat{y} - x\|^2 + \frac{1}{\alpha}\|\tilde{y} - \hat{y}\| \cdot \|\hat{y} - x\| \\
&\leq \left(1 + \frac{\tau}{2}\right)(-h(\tilde{y}; x, s)) + \frac{\tau}{2}(-h(\tilde{y}; x, s)) + 2\sqrt{1 + \frac{\tau}{2}}\sqrt{\frac{\tau}{2}}(-h(\tilde{y}; x, s)) \\
&= \left(\sqrt{1 + \frac{\tau}{2}} + \sqrt{\frac{\tau}{2}}\right)^2 (-h(\tilde{y}; x, s)),
\end{aligned}$$

where the last inequality follows from the application of (4.11)-(4.12). ∎

The next lemma provides a subgradient $\hat{v} \in \partial f(\hat{y})$ whose norm is bounded from above by a quantity containing $\sqrt{-h(\tilde{y}; x, s)}$. Its proof is omitted since it is almost identical to the one of Lemma 3 in [16].

LEMMA 4.2. *Suppose Assumptions [A1]–[A3] hold true. Let x be a point in $\mathrm{dom}(f_1)$ and let $\hat{y}, \tilde{y}$ be defined as in (4.6)–(4.8). Moreover, assume that $\alpha \in [\alpha_{min}, \alpha_{max}]$, with $0 < \alpha_{min} \leq \alpha_{max}$ and $\beta \in [0, \beta_{max}]$, with $\beta_{max} \geq 0$. Then, there exists a subgradient $\hat{v} \in \partial f(\hat{y})$ such that*

$$\|\hat{v}\| \leq p(\|\hat{y} - x\| + \|x - s\|) \tag{4.14}$$

$$\leq q(\sqrt{-h(\tilde{y}; x, s)} + \|x - s\|), \tag{4.15}$$

where the two constants p, q depend only on $\alpha_{min}, \alpha_{max}, \beta_{max}$ and on the Lipschitz constant L.

The following lemma is the equivalent of Lemma 4-5 in [16].

LEMMA 4.3. *Suppose Assumptions [A1]–[A3] hold true and assume $0 < \alpha_{min} \leq \alpha \leq \alpha_{max}$, $\beta \in [0, \beta_{max}]$. Let (x, s) be a point in $\mathrm{dom}(f_1) \times \mathbb{R}^n$ and let $\hat{y}, \tilde{y}$ be defined as in (4.6)–(4.8), for some $\tau \geq 0$. Then, there exists $c, d, \bar{c}, \bar{d} \in \mathbb{R}$ depending only on $\alpha_{min}, \alpha_{max}, \beta_{max}, \tau$ such that*

$$f(\hat{y}) \geq f(\tilde{y}) + ch(\tilde{y}; x, s) - d\|x - s\|^2 \tag{4.16}$$

$$f(\hat{y}) \leq f(x) - \bar{c}h(\tilde{y}; x, s) + \bar{d}\|x - s\|^2. \tag{4.17}$$

Proof. From the Descent Lemma [5, Proposition A.24] we have

$$f_0(\hat{y}) \geq f_0(\tilde{y}) - \langle \nabla f_0(\hat{y}), \tilde{y} - \hat{y} \rangle - \frac{L}{2}\|\tilde{y} - \hat{y}\|^2. \tag{4.18}$$

The inclusion $0 \in \partial_\epsilon h(\tilde{y}; x, s)$ in (4.10) implies that there exists a vector $e \in \mathbb{R}^n$ with

$$\frac{1}{2\alpha}\|e\|^2 \leq \epsilon \tag{4.19}$$

such that

$$-\frac{1}{\alpha}(\tilde{y} - x + \alpha\nabla f_0(x) - \beta(x - s) + e) \in \partial_\epsilon f_1(\tilde{y})$$

(see [12] and references therein). The definition of ϵ-subdifferential implies

$$f_1(\hat{y}) \geq f_1(\tilde{y}) - \frac{1}{\alpha}\langle \tilde{y} - x, \hat{y} - \tilde{y} \rangle + \frac{\beta}{\alpha}\langle \hat{y} - \tilde{y}, x - s \rangle - \langle \nabla f_0(x), \hat{y} - \tilde{y} \rangle - \frac{1}{\alpha}\langle e, \hat{y} - \tilde{y} \rangle - \epsilon. \tag{4.20}$$

Summing inequalities (4.18) and (4.20) yields

$$\begin{aligned} f(\hat{y}) \geq f(\tilde{y}) &- \langle \nabla f_0(x) - \nabla f_0(\hat{y}), \hat{y} - \tilde{y} \rangle + \frac{\beta}{\alpha}\langle \hat{y} - \tilde{y}, x - s \rangle \\ &- \frac{1}{\alpha}\langle \tilde{y} - x, \hat{y} - \tilde{y} \rangle - \frac{1}{\alpha}\langle e, \hat{y} - \tilde{y} \rangle - \frac{L}{2}\|\tilde{y} - \hat{y}\|^2 - \epsilon. \end{aligned} \tag{4.21}$$

Now we consider each term at the right-hand-side in the above inequality so as to obtain a lower bound. Using the Cauchy-Schwarz inequality, Assumption [A3], (4.11) and (4.12) we obtain

$$\begin{aligned} \langle \nabla f_0(x) - \nabla f_0(\hat{y}), \hat{y} - \tilde{y} \rangle &\leq \|\nabla f_0(x) - \nabla f_0(\hat{y})\|\|\hat{y} - \tilde{y}\| \\ &\leq L\alpha_{max}\sqrt{2\tau\left(1 + \frac{\tau}{2}\right)}(-h(\tilde{y}; x, s)). \end{aligned} \tag{4.22}$$

Similarly, using again the Cauchy-Schwarz inequality, (4.12) and (4.13), we can write

$$\frac{1}{\alpha}\langle \tilde{y} - x, \hat{y} - \tilde{y} \rangle \leq \frac{1}{\alpha_{min}}\|\tilde{y} - x\|\|\hat{y} - \tilde{y}\| \leq \frac{\alpha_{max}}{\alpha_{min}}\sqrt{\frac{2\tau}{\theta}}(-h(\tilde{y}; x, s)). \tag{4.23}$$

Moreover, from (4.19) and (4.10) we obtain $\|e\| \leq \sqrt{2\alpha_{max}\epsilon} \leq \sqrt{\alpha_{max}\tau(-h(\tilde{y};x,s))}$ which, using also (4.12), yields

$$\frac{1}{\alpha}\langle e, \hat{y}-\tilde{y}\rangle \leq \frac{1}{\alpha}\|e\|\|\hat{y}-\tilde{y}\| \leq \frac{\alpha_{max}\tau}{\alpha_{min}}(-h(\tilde{y};x,s)). \tag{4.24}$$

Finally, using (4.12), we can also write

$$\frac{\beta}{\alpha}\langle \hat{y}-\tilde{y}, x-s\rangle \geq -\frac{\beta}{2\alpha}(\|\hat{y}-\tilde{y}\|^2+\|x-s\|^2) \geq -\frac{\beta_{max}}{2\alpha_{min}}\left(-\alpha_{max}\tau h(\tilde{y};x,s)+\|x-s\|^2\right). \tag{4.25}$$

Combining (4.21) with (4.13), (4.22), (4.23), (4.24), (4.25) and (4.10), gives (4.16) with

$$c = L\alpha_{max}\sqrt{2\tau\left(1+\frac{\tau}{2}\right)} + \frac{\alpha_{max}}{\alpha_{min}}\sqrt{\frac{2\tau}{\theta}} + \frac{\alpha_{max}\tau}{\alpha_{min}} + \frac{L\alpha_{max}\tau}{2} + \frac{\tau}{2} + \frac{\beta_{max}\alpha_{max}\tau}{2\alpha_{min}}, \quad d = \frac{\beta_{max}}{2\alpha_{min}}.$$

As for (4.17), using the Descent Lemma we obtain

$$f_0(y) \leq f_0(x) + \langle \nabla f_0(x), y-x\rangle + \frac{L}{2}\|y-x\|^2,$$

for all $x, y \in \mathrm{dom}(f_1)$. Summing $f_1(y)$ on both sides yields

$$\begin{aligned} f(y) &\leq f(x) + f_1(y) - f_1(x) + \langle \nabla f_0(x), y-x\rangle + \frac{L}{2}\|y-x\|^2 \\ &\leq f(x) + h(y;x,s) + \frac{L}{2}\|y-x\|^2 + \frac{\beta}{\alpha}\langle x-s, y-x\rangle \\ &\leq f(x) + h(y;x,s) + \frac{L}{2}\|y-x\|^2 + \frac{\beta}{2\alpha}(\|x-s\|^2+\|y-x\|^2). \end{aligned}$$

From the previous inequality with $y=\hat{y}$, recalling that $h(\hat{y};x,s)\leq 0$ and combining with (4.11) yields (4.17), where the constants are set as $\bar{c} = (L\alpha_{max}+\beta_{max})(1+\tau/2)$, $\bar{d} = \beta_{max}/(2\alpha_{min})$. □

4.2. i²Piano: inertial inexact Proximal algorithm for nonconvex optimization. In this section we propose a generalization of the inertial method in [30], introducing the possibility of an inexact computation of the inertial proximal gradient point. We will refer to the new algorithm as i²Piano (inertial inexact Proximal algorithm for nonconvex optimization). Let us describe the i²Piano iteration as follows. STEP 1–4 determine the stepsize α_k and inertial parameter β_k at iteration k. Given the parameters α_k, β_k, STEP 5 seeks to find a possibly inexact inertial proximal point, i.e., an inexact minimizer of the function

$$h^{(k)}(y;x,s) = f_1(y) - f_1(x) + \langle \nabla f_0(x) - \frac{\beta_k}{\alpha_k}(x-s), y-x\rangle + \frac{1}{2\alpha_k}\|y-x\|^2. \tag{4.26}$$

According to (4.8)–(4.10), the i²Piano iterate is any point $x^{(k+1)} = \tilde{y}^{(k)}$ such that

$$0 \in \partial_{\epsilon_k} h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}), \quad \text{with } \epsilon_k = -\frac{\tau}{2}h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}) \tag{4.27}$$

for some fixed nonnegative constant τ. When $\tau=0$, we recover the exact inertial proximal gradient point provided by the iPiano method [30, 31]. As explained at the beginning of Section 4, STEP 5 of i²Piano can be practically implemented with an inner loop consisting of an iterative optimization method applied to the dual of problem $\min_{y\in\mathbb{R}^n} h^{(k)}(y;x^{(k)},x^{(k-1)})$, until the duality gap is smaller than ϵ_k. In the implementation of i²Piano, besides the stepsize α_k and the inertial parameter β_k, a further parameter, L_k, is introduced (cf. STEP 6) with the aim of estimating a local Lipschitz constant of ∇f_0 that allows us to take larger steps. In particular, L_k is successively increased by a factor $\eta > 1$ until the following descent condition holds

$$f_0(x^{(k+1)}) \leq f_0(x^{(k)}) + \langle \nabla f_0(x^{(k)}), x^{(k+1)}-x^{(k)}\rangle + \frac{L_k}{2}\|x^{(k+1)}-x^{(k)}\|^2. \tag{4.28}$$

i^{2}Piano inertial inexact Proximal algorithm for nonconvex optimization.

Choose $x^{(-1)}, x^{(0)} \in \mathrm{dom}(f_1)$, $\delta \geq \gamma > 0$, $\eta > 1$, $0 < L_{min} \leq L_{max}$, $\tau \geq 0$. Set $\theta = 2/(\sqrt{2+\tau} + \sqrt{\tau})^2$ and choose $\omega \in [0,1)$ if $\tau > 0$, $\omega \in [0,1]$ if $\tau = 0$.
FOR $k = 0, 1, \ldots$

STEP 1. Choose $L_k \in [L_{min}, L_{max}]$.

STEP 2. Set $b_k = \dfrac{L_k + 2\delta}{L_k + 2\gamma}$.

STEP 3. Set $\beta_k = \dfrac{1+\theta\omega}{2} \cdot \dfrac{b_k - 1}{b_k - \frac{1}{2}}$.

STEP 4. Set $\alpha_k = \dfrac{1 + \theta\omega - 2\beta_k}{L_k + 2\gamma}$.

STEP 5. Compute $\tilde{y}^{(k)}$ such that

$$0 \in \partial_{\epsilon_k} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, x^{(k-1)}), \qquad \text{with } \epsilon_k = -\frac{\tau}{2} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, x^{(k-1)})$$

STEP 6. Check the local descent:

IF $f_0(\tilde{y}^{(k)}) \leq f_0(x^{(k)}) + \langle \nabla f_0(x^{(k)}), \tilde{y}^{(k)} - x^{(k)} \rangle + \dfrac{L_k}{2} \|\tilde{y}^{(k)} - x^{(k)}\|^2$

- Set $x^{(k+1)} = \tilde{y}^{(k)}$.

ELSE

- Set $L_k = \eta L_k$.
- Go to STEP 2.

END

4.2.1. Convergence analysis. Our aim now is to frame i^{2}Piano in the abstract scheme defined by Conditions 3.1 to enjoy the favourable convergence guarantees that are provided by Theorem 3.3. We start the convergence analysis by showing that i^{2}Piano is well-posed and that its parameters satisfy some useful relations.

LEMMA 4.4. *The loop between* STEP 2 *and* STEP 6 *terminates in a finite number of steps. In particular, there exists* $\overline{L} > 0$ *such that* $L_k \leq \overline{L}$, $\forall\ k \geq 0$. *Moreover, we have*

$$0 \leq \beta_k \leq \frac{1+\theta\omega}{2}, \quad \forall\ k \geq 0 \tag{4.29}$$

and there exist two positive constants $\alpha_{min}, \alpha_{max}$ *with* $0 < \alpha_{min} \leq \alpha_{max}$ *such that* $\alpha_k \in [\alpha_{min}, \alpha_{max}]$, $\forall k \geq 0$. *We also have*

$$\frac{1+\theta\omega}{2\alpha_k} - \frac{L_k}{2} - \frac{\beta_k}{2\alpha_k} = \delta \tag{4.30}$$

$$\delta - \frac{\beta_k}{2\alpha_k} = \gamma. \tag{4.31}$$

Proof. Since $\eta > 1$, after a finite number of steps the tentative value of L_k satisfies $L_k \geq L$, where L is the Lipschitz constant of ∇f_0. Then, from the Descent Lemma, the inequality at STEP 6 is satisfied. From $\delta \geq \gamma$ we have $b_k \geq 1$, which implies (4.29). A simple inspection shows that the following equalities hold:

$$b_k = \frac{1 + \theta\omega - \beta_k}{1 + \theta\omega - 2\beta_k} \Rightarrow \frac{L_k + 2\delta}{L_k + 2\gamma} = \frac{1 + \theta\omega - \beta_k}{1 + \theta\omega - 2\beta_k},$$

which leads to rewriting the parameter α_k as

$$\alpha_k = \frac{1 + \theta\omega - \beta_k}{L_k + 2\delta}. \tag{4.32}$$

Then there holds $\alpha_k \geq \alpha_{min}$ with $\alpha_{min} = (1+\theta\omega)/(2(\overline{L} + 2\delta))$ and, since $\theta\omega \leq 1$, we also have $\alpha_k \leq \alpha_{max}$ with $\alpha_{max} = 2/L_{min}$. Moreover, we have

$$\frac{1+\theta\omega}{\alpha_k} - \frac{L_k}{2} - \frac{\beta_k}{2\alpha_k} = \frac{1 + \theta\omega - \beta_k}{2\alpha_k} - \frac{L_k}{2} = \frac{L_k + 2\delta}{2} - \frac{L_k}{2} = \delta$$

and

$$\delta - \frac{\beta_k}{2\alpha_k} = \frac{1+\theta\omega}{2\alpha_k} - \frac{L_k}{2} - \frac{\beta_k}{\alpha_k} = \frac{1+\theta\omega-2\beta_k}{2\alpha_k} - \frac{L_k}{2} = \frac{L_k+2\gamma}{2} - \frac{L_k}{2} = \gamma.$$

□

Notice that the case $\tau = 0$, which corresponds to the exact computation of the inertial proximal gradient point at STEP 5, implies $\theta = 1$ and, choosing $\omega = 1$, the parameters settings in i^{2}Piano are exactly the same as in [31]. The need of introducing the parameter ω is mainly technical and will be explained in the following.
We now prove that condition *[H1]* holds for i^{2}Piano when the corresponding surrogate function Φ is defined as follows:

$$\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \qquad \Phi(x,s) = f(x) + \delta\|x-s\|^2. \tag{4.33}$$

PROPOSITION 4.5. *Let $\{x^{(k)}\}_{k\in\mathbb{N}}$ be the sequence generated by i^{2}Piano. Then there holds*

$$\Phi(x^{(k+1)},x^{(k)}) \le \Phi(x^{(k)},x^{(k-1)}) - \gamma\|x^{(k)}-x^{(k-1)}\|^2 + (1-\omega)h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}). \tag{4.34}$$

Consequently, there exist $\{s^{(k)}\}_{k\in\mathbb{N}}$, $\{d_k\}_{k\in\mathbb{N}}$, $\{a_k\}_{k\in\mathbb{N}}$ such that [H1] holds with Φ as in (4.33).

Proof. By summing the quantity $f_1(x^{(k+1)})$ to both sides of inequality (4.28) we obtain

$$\begin{aligned}
f(x^{(k+1)}) &\le f(x^{(k)}) + f_1(x^{(k+1)}) - f_1(x^{(k)}) + \langle \nabla f_0(x^{(k)}), x^{(k+1)}-x^{(k)}\rangle + \frac{L_k}{2}\|x^{(k+1)}-x^{(k)}\|^2 \\
&= f(x^{(k)}) + h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}) - \left(\frac{1}{2\alpha_k} - \frac{L_k}{2}\right)\|x^{(k+1)}-x^{(k)}\|^2 \\
&\quad + \frac{\beta_k}{\alpha_k}\langle x^{(k+1)}-x^{(k)}, x^{(k)}-x^{(k-1)}\rangle \\
&\le f(x^{(k)}) + h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}) - \left(\frac{1}{2\alpha_k} - \frac{L_k}{2}\right)\|x^{(k+1)}-x^{(k)}\|^2 \\
&\quad + \frac{\beta_k}{2\alpha_k}(\|x^{(k+1)}-x^{(k)}\|^2 + \|x^{(k)}-x^{(k-1)}\|^2) \\
&\le f(x^{(k)}) + (1-\omega)h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}) - \frac{\theta\omega}{2\alpha_k}\|x^{(k+1)}-x^{(k)}\|^2 \\
&\quad - \left(\frac{1}{2\alpha_k} - \frac{L_k}{2}\right)\|x^{(k+1)}-x^{(k)}\|^2 + \frac{\beta_k}{2\alpha_k}(\|x^{(k+1)}-x^{(k)}\|^2 + \|x^{(k)}-x^{(k-1)}\|^2) \\
&= f(x^{(k)}) - \left(\frac{1+\theta\omega}{2\alpha_k} - \frac{L_k}{2} - \frac{\beta_k}{2\alpha_k}\right)\|x^{(k+1)}-x^{(k)}\|^2 \\
&\quad + (1-\omega)h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}) + \frac{\beta_k}{2\alpha_k}\|x^{(k)}-x^{(k-1)}\|^2
\end{aligned}$$

where the first equality is obtained by adding and subtracting to the right-hand-side the quantity $\|x^{(k+1)}-x^{(k)}\|^2/(2\alpha_k) + \beta_k/\alpha_k\langle x^{(k+1)}-x^{(k)}, x^{(k)}-x^{(k-1)}\rangle$, the subsequent inequality follows from the basic relation $2\langle a,b\rangle \le \|a\|^2 + \|b\|^2$ and the next one from (4.13).
Recalling (4.30), the above inequality can be conveniently rewritten as

$$\begin{aligned}
f(x^{(k+1)}) + \delta\|x^{(k+1)}-x^{(k)}\|^2 &\le f(x^{(k)}) + \delta\|x^{(k)}-x^{(k-1)}\|^2 \\
&\quad + \left(\frac{\beta_k}{2\alpha_k} - \delta\right)\|x^{(k)}-x^{(k-1)}\|^2 + (1-\omega)h(x^{(k+1)};x^{(k)},x^{(k-1)}).
\end{aligned}$$

Finally, exploiting (4.31), we obtain condition *[H1]* with Φ given in (4.33), $a_k = 1$ and

$$s^{(k)} = x^{(k-1)} \tag{4.35}$$

$$d_k^2 = \gamma\|x^{(k)}-x^{(k-1)}\|^2 - (1-\omega)h^{(k)}(x^{(k+1)};x^{(k)},x^{(k-1)}). \tag{4.36}$$

□

Under Assumption *[A4]*, Φ is bounded from below, hence, condition *[H1]* implies

$$\lim_{k\to\infty} \|x^{(k)} - x^{(k-1)}\| = 0 \tag{4.37}$$

and, if $\omega < 1$, also

$$\lim_{k\to\infty} h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)}) = 0. \tag{4.38}$$

The choice $\omega < 1$ is enforced when $\tau > 0$ in order to obtain (4.38). If we take $\omega = 1$, with the same arguments as above we still obtain (4.37). It is also worth noticing that, for large values of τ, that is when a coarser accuracy is allowed in the computation of $\tilde{y}^{(k)}$, the parameter θ can be very small and this also influences the choice of α_k and β_k.

In order to prove condition *[H2]*, let us now introduce the second surrogate function as follows

$$\mathcal{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, \quad \mathcal{F}(u, \rho) = f(u) + \frac{1}{2}\rho^2. \tag{4.39}$$

The following result is proved using similar arguments as the ones used in Lemma 4-5 in [16].

PROPOSITION 4.6. *Let $\{x^{(k)}\}_{k\in\mathbb{N}}$ be the sequence generated by i^{2}Piano and, for each k, let the point $\hat{x}^{(k+1)}$ be defined as*

$$\hat{x}^{(k+1)} = \operatorname*{argmin}_{y\in\mathbb{R}^n} h^{(k)}(y; x^{(k)}, x^{(k-1)}). \tag{4.40}$$

Then, there exist $\{\rho^{(k)}\}_{k\in\mathbb{N}}$, $\{r_k\}_{k\in\mathbb{N}}$ such that condition [H2] holds with Φ defined in (4.33), $\mathcal{F}$ defined in (4.39), $s^{(k)}$ defined in (4.35) and $u^{(k)} = \hat{x}^{(k+1)}$.

Proof. When $\omega = 1$ we necessarily have $\tau = 0$, which means $x^{(k+1)} = \hat{x}^{(k+1)}$. Therefore, *[H2]* direcly follows from *[H1]* with $u^{(k)} = x^{(k+1)}$, $\rho^{(k)} = \sqrt{2\delta}\|x^{(k)} - x^{(k-1)}\|$, $r_k = 0$. Consider now the case $\omega < 1$. From (4.16) and (4.17), we directly obtain

$$\begin{aligned} f(\hat{x}^{(k+1)}) &\geq f(x^{(k+1)}) + ch^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)}) - d\|x^{(k)} - x^{(k-1)}\|^2 \\ f(\hat{x}^{(k+1)}) &\leq f(x^{(k)}) - \bar{c}h(x^{(k+1)}; x^{(k)}, x^{(k-1)}) + \bar{d}\|x^{(k)} - x^{(k-1)}\|^2, \end{aligned}$$

where $c, d, \bar{c}, \bar{d}$ are defined as in Lemma 4.3 and do not depend on k. Combining the two inequalities above we obtain

$$\begin{aligned} &f(x^{(k+1)}) + \delta\|x^{(k+1)} - x^{(k)}\|^2 \leq \\ &\quad \leq f(\hat{x}^{(k+1)}) + \delta\|x^{(k+1)} - x^{(k)}\|^2 + c(-h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)})) + d\|x^{(k)} - x^{(k-1)}\|^2 \\ &\quad \leq f(x^{(k)}) - (c + \bar{c})h(x^{(k+1)}; x^{(k)}, x^{(k-1)}) + (d + \bar{d})\|x^{(k)} - x^{(k-1)}\|^2 + \delta\|x^{(k+1)} - x^{(k)}\|^2. \end{aligned}$$

Recalling the definition of $\mathcal{F}$ in (4.39), the above inequalities can be rewritten as

$$f(x^{(k+1)}) + \delta\|x^{(k+1)} - x^{(k)}\|^2 \leq \mathcal{F}(\hat{x}^{(k+1)}, \rho^{(k)}) \leq f(x^{(k)}) + \delta\|x^{(k)} - x^{(k-1)}\|^2 + r_k \tag{4.41}$$

where $\rho^{(k)}, r_k$ are given by

$$\begin{aligned} \rho^{(k)} &= \sqrt{2}(\delta\|x^{(k+1)} - x^{(k)}\|^2 + c(-h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)})) + d\|x^{(k)} - x^{(k-1)}\|^2)^{\frac{1}{2}} \\ r_k &= -(c + \bar{c})h(x^{(k+1)}; x^{(k)}, x^{(k-1)}) + (d + \bar{d})\|x^{(k)} - x^{(k-1)}\|^2 + \delta\|x^{(k+1)} - x^{(k)}\|^2. \end{aligned} \tag{4.42}$$

From (4.37)–(4.38) we obtain $\lim_{k\to\infty} r_k = 0$, hence, assumption *[H2]* is satisfied with the above settings and with $u^{(k)} = \hat{x}^{(k+1)}$. □

Next we show that condition *[H3]* holds for i^{2}Piano. The following result combines elements of Lemma 3 in [16] and Lemma 17 in [30].

PROPOSITION 4.7. *There exist $b > 0$, $I \subset \mathbb{Z}$, $\{\theta_i\}_{i\in I}$ such that condition [H3] holds with $\{u^{(k)}\}_{k\in\mathbb{N}}, \{\rho^{(k)}\}_{k\in\mathbb{N}}, \{d_k\}_{k\in\mathbb{N}}$ defined as in Proposition 4.6 and in (4.36), $\zeta_{k+1} = 0$ and $b_{k+1} = 1$.*

Proof. From the separable structure of $\mathcal{F}$, if $v \in \partial f(u)$ and $\rho \in \mathbb{R}$, we have that $(v, \rho) \in \partial\mathcal{F}(u, \rho)$. In particular,

$$\|\partial\mathcal{F}(u,\rho)\|_- \leq \|v\| + |\rho|, \text{ for all } v \in \partial f(u), \rho \in \mathbb{R}. \tag{4.43}$$

If $\omega = 1$, we are in the case $\tau = 0$. This means that $x^{(k+1)} = \hat{x}^{(k+1)}$, $\rho^{(k)} = \sqrt{2\delta}\|x^{(k)} - x^{(k-1)}\|$. Moreover, (4.36) reduces to $d_k = \sqrt{\gamma}\|x^{(k)} - x^{(k-1)}\| = \sqrt{\gamma/(2\delta)}\rho^{(k)}$. From (4.14), we have that there exists a subgradient $\hat{v}^{(k+1)} \in \partial f(\hat{x}^{(k+1)})$ and a positive constant p such that

$$\|\hat{v}^{(k+1)}\| \leq p(\|x^{(k+1)} - x^{(k)}\| + \|x^{(k)} - x^{(k-1)}\|) = \frac{p}{\sqrt{\gamma}}(d_{k+1} + d_k).$$

Hence, $\|\hat{v}^{(k+1)}\| + \rho^{(k)} \leq pd_{k+1}/\sqrt{\gamma} + d_k(p + \sqrt{2\delta})/\sqrt{\gamma}$ and, recalling (4.43), *[H3]* follows with $b = (p + \sqrt{2\delta})/\sqrt{\gamma}$, $I = \{0, 1\}$, $\theta_0 = \theta_1 = \frac{1}{2}$.
Consider now the case $\omega < 1$. From (4.15), there exists a subgradient $\hat{v}^{(k+1)} \in \partial f(\hat{x}^{(k+1)})$ and a positive constant q such that

$$\begin{aligned}
\|\hat{v}^{(k+1)}\| &\leq q\sqrt{-h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)})} + q\|x^{(k)} - x^{(k-1)}\| \\
&\leq q\sqrt{-h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)})} + q\sqrt{\frac{2\alpha_{max}}{\theta}}\sqrt{-h^{(k-1)}(x^{(k)}; x^{(k-1)}, x^{(k-2)})} \\
&\leq \frac{q}{\sqrt{1-\omega}}\sqrt{-(1-\omega)h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)}) + \gamma\|x^{(k)} - x^{(k-1)}\|^2} \\
&\quad + q\sqrt{\frac{2\alpha_{max}}{\theta(1-\omega)}}\sqrt{-(1-\omega)h^{(k-1)}(x^{(k)}; x^{(k-1)}, x^{(k-2)}) + \gamma\|x^{(k-1)} - x^{(k-2)}\|^2} \qquad (4.44) \\
&\leq \frac{q}{\sqrt{1-\omega}}d_k + q\sqrt{\frac{2\alpha_{max}}{\theta(1-\omega)}}d_{k-1}. \qquad (4.45)
\end{aligned}$$

From (4.42) and (4.13) we have

$$\rho^{(k)} \leq \sqrt{2}\left(-(2\delta\alpha_{max}/\theta + c)h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)}) + d\|x^{(k)} - x^{(k-1)}\|^2\right)^{\frac{1}{2}} \leq r \cdot d_k \tag{4.46}$$

where $r = \sqrt{2}\max\{(2\delta\alpha_{max}/\theta + c)/(1-\omega), d/\gamma\}^{\frac{1}{2}}$. Then, combining (4.45) with (4.46), in view of (4.43) we obtain

$$\|\partial\mathcal{F}(\hat{x}^{(k+1)}, \rho^{(k)})\|_- \leq \frac{b}{2}(d_k + d_{k-1}), \quad \text{with } b = 2\max\left\{\frac{q}{\sqrt{1-\omega}} + r, q\sqrt{\frac{2\alpha_{max}}{\theta(1-\omega)}}\right\}$$

and the thesis follows with $I = \{1, 2\}$, $\theta_1 = \theta_2 = \frac{1}{2}$. □

The following lemma holds for all methods whose iterates satisfy *[H1]*,*[H2]*, *[H3]* with $\mathcal{F}$ defined as in (4.39), and it is crucial to ensure condition *[H4]* for i^{2}Piano.

PROPOSITION 4.8. *Let Assumptions [A1]–[A4] be satisfied and assume that $\{x^{(k)}\}_{k\in\mathbb{N}}$ satisfies [H1],[H2], [H3] with $\mathcal{F}$ defined as in (4.39). If $\{(x^{(k_j)}, \rho^{(k_j)})\}_{j\in\mathbb{N}}$ is a subsequence of $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ converging to some $(x^*, \rho^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $\lim_{j\to\infty}\|u^{(k_j)} - x^{(k_j)}\| = 0$, then $\rho^* = 0$ and $\lim_{j\to\infty}\mathcal{F}(u^{(k_j)}, \rho^{(k_j)}) = \mathcal{F}(x^*, \rho^*)$.*

Proof. Recalling that *[H1]* implies $\lim_{k\to\infty} d_k = 0$, from *[H3]* and thanks to the separable structure of $\mathcal{F}$, we obtain that there exists $\hat{v}^{(k)} \in \partial f(u^{(k)})$ such that $\lim_{k\to\infty}\|\hat{v}^{(k)}\| = \lim_{k\to\infty}\rho^{(k)} = 0$. In particular, in view of (4.2), we can write $\hat{v}^{(k)} = \nabla f_0(u^{(k)}) + w^{(k)}$, where $w^{(k)} \in \partial f_1(u^{(k)})$. Therefore, by continuity of ∇f_0, the following implication holds

$$\lim_{k\to\infty}\hat{v}^{(k)} = 0 \Rightarrow \lim_{j\to\infty} w^{(k_j)} = -\nabla f_0(x^*).$$

Adding the quantity $\frac{1}{2}(\rho^{(k_j)})^2$ to both sides of the subgradient inequality yields

$$\begin{aligned} f_1(x^*) + \frac{1}{2}(\rho^{(k_j)})^2 &\geq f_1(u^{(k_j)}) + \langle w^{(k_j)}, x^* - u^{(k_j)} \rangle + \frac{1}{2}(\rho^{(k_j)})^2 \\ &= \mathcal{F}(u^{(k_j)}, \rho^{(k_j)}) - f_0(u^{(k_j)}) + \langle w^{(k_j)}, x^* - u^{(k_j)} \rangle \end{aligned}$$

where the last equality is obtained by adding and subtracting $f_0(u^{(k_j)})$ to the right-hand-side. Taking limits on both sides we obtain

$$f_1(x^*) + \frac{1}{2}(\rho^*)^2 \geq \lim_{j\to\infty} \mathcal{F}(u^{(k_j)}, \rho^{(k_j)}) - f_0(x^*),$$

which, rearranging terms, gives $\lim_{j\to\infty} \mathcal{F}(u^{(k_j)}, \rho^{(k_j)}) \leq \mathcal{F}(x^*, \rho^*)$. On the other side, by assumption, $\mathcal{F}$ is lower semicontinuous, therefore $\lim_{j\to\infty} \mathcal{F}(u^{(k_j)}, \rho^{(k_j)}) \geq \mathcal{F}(x^*, \rho^*)$, which completes the proof. □

We are now ready to prove that i^{2}Piano complies with Conditions 3.1 and, hence, its iterates converge to a stationary point.

THEOREM 4.9. *Suppose that $\mathcal{F}$ is a KL function and assume that the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ generated by i^{2}Piano is bounded. Then, $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to a stationary point of f.*

Proof. By Propositions 4.5-4.6-4.7, we know that conditions *[H1]-[H2]-[H3]* hold for i^{2}Piano. Furthermore, if $\omega = 1$, then $u^{(k)} = \hat{x}^{(k+1)} = x^{(k+1)}$, and (4.37) directly implies that $\lim_{k\to\infty} \|u^{(k)} - x^{(k)}\| = 0$. If $\omega < 1$, using (4.11) and (4.38) we have

$$\lim_{k\to\infty} \|u^{(k)} - x^{(k)}\| = \lim_{k\to\infty} \|\hat{x}^{(k+1)} - x^{(k)}\| \leq \lim_{k\to\infty} \sqrt{-2\alpha_{max}\left(1 + \frac{\tau}{2}\right) h^{(k)}(x^{(k+1)}; x^{(k)}, x^{(k-1)})} = 0.$$

Then, in both cases, the assumptions of Proposition 4.8 are satisfied and, therefore, condition *[H4]* holds. Finally, condition *[H5]* follows from (4.36), while condition *[H6]* is trivially satisfied, since both sequences $\{a_k\}_{k\in\mathbb{N}}$ and $\{b_k\}_{k\in\mathbb{N}}$ are constant. Then, Theorem 3.3 applies and guarantees that the sequence $\{(x^{(k)}, \rho^{(k)})\}_{k\in\mathbb{N}}$ converges to a stationary point (x^*, ρ^*) of $\mathcal{F}$. Note that, since $\mathcal{F}$ is the sum of separable functions, its subdifferential can be written as $\partial\mathcal{F}(x, \rho) = \partial f(x) \times \{\rho\}$. Then, (x^*, ρ^*) is stationary for $\mathcal{F}$ if and only if $\rho^* = 0$ and $0 \in \partial f(x^*)$. Hence, x^* is a stationary point for f and $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to it. □

Remark 4.10. We underline that $\mathcal{F}$ is a KL function if, for instance, f and $\frac{1}{2}\|\cdot\|^2$ are definable in the same $o-$minimal structure [7, Definition 7]. Indeed functions definable in an $o-$minimal structure satisfy the KL property on their domain [7, Theorem 11] and $o-$minimal structures are closed with respect to the sum, see [7, Remark 5] and references therein. Examples of functions definable in an $o-$minimal structure are semialgebraic, subanalytic and real analytic functions.

Remark 4.11. Theorem 4.9 requires the boundedness of the iterates as hypothesis. A standard way to assert such a condition is when the Lyapunov function Φ defined in (4.33) is coercive, since this assumption combined with the descent property (4.34) guarantees that the sequence $\{(x^{(k)}, x^{(k-1)})\}_{k\in\mathbb{N}}$ is included in a (bounded) level set of the coercive function Φ.

4.3. iPila: inertial Proximal inexact line–search algorithm. In the following we introduce a novel algorithm combining a line–search along the descent direction and an inertial proximal-gradient step as a special case of our abstract scheme. A line–search procedure for the objective function f requires a descent direction $d \in \mathbb{R}^n$, i.e., a vector such that the directional derivative $f'(x; d) = \lim_{\lambda\downarrow 0}(f(x+\lambda) - f(x))/\lambda$ is negative. As explained extensively in [11, 12], the inexact proximal-gradient point provides a descent direction for the objective function f. Indeed, if $\tilde{y}$ satisfies (4.8), where the inertial parameter β in (4.5) is equal to zero, then the vector $\tilde{y} - x$ is a descent direction for f at x. Unfortunately, this is not true, in general, when $\beta > 0$. In this section we show that in the general case $\beta \geq 0$, the point $\tilde{y}$ can be still used to define a descent direction for a suitable merit function. Then, we propose a line–search procedure along this direction that enable us to define a descent algorithm such that the merit function monotonically

decreases along the iterates. Differently from the backtracking procedure in i^{2}Piano, the proposed line–search requires to solve the minimization subproblem (4.4) only once per iteration. Finally, we show that the new algorithm can be analyzed in the framework of Section 3, in order to prove the convergence of the iterates to a stationary point of f. In this case, unlike i^{2}Piano, one of the two merit functions involved in the abstract scheme defined by Conditions 3.1 will play an active role in the algorithm, as it will be explicitly computed at each iteration to determine the new point. In particular, we define the merit function Φ appearing in *[H1]–[H2]* as follows:

$$\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad \Phi(x,s) = f(x) + \frac{1}{2}\|x-s\|^2,$$

where the variable s will be considered as an actual optimization variable, independent on x. The composite structure of Φ is $\Phi(x,s) = \Phi_0(x,s) + \Phi_1(x,s)$, with

$$\Phi_0(x,s) = f_0(x) + \frac{1}{2}\|x-s\|^2, \quad \Phi_1(x,s) = f_1(x).$$

The function Φ_0 is differentiable with gradient

$$\nabla\Phi_0(x,s) = \begin{pmatrix} \nabla_x\Phi_0(x,s) \\ \nabla_s\Phi_0(x,s) \end{pmatrix} = \begin{pmatrix} \nabla f_0(x) + x - s \\ s - x \end{pmatrix}.$$

It is easy to see that $\nabla\Phi_0(x,s)$ is Lipschitz continuous and in particular it holds that

$$\|\nabla\Phi_0(x,s) - \nabla\Phi_0(\bar{x},\bar{s})\| \le M \left\| \begin{pmatrix} x-\bar{x} \\ s-\bar{s} \end{pmatrix} \right\|, \quad \forall x,\bar{x},s,\bar{s}, \tag{4.47}$$

with $M = L + 2$, where L is the Lipschitz constant of ∇f_0.
Given a vector $d \in \mathbb{R}^{2n}$

$$d = \begin{pmatrix} d_x \\ d_s \end{pmatrix}, \tag{4.48}$$

the directional derivative of Φ at the point (x,s) with respect to the direction d can be written as

$$\begin{aligned} \Phi'(x,s;d_x,d_s) &= \Phi_0'(x,s;d_x,d_s) + \Phi_1'(x,s;d_x,d_s) = f_1'(x;d_x) + \langle \nabla\Phi_0(x,s), d\rangle \\ &= f_1'(x;d_x) + \langle \nabla f_0(x) + x - s, d_x\rangle + \langle s - x, d_s\rangle, \end{aligned}$$

which always exists thanks to the convexity of f_1.
A vector $d \in \mathbb{R}^{2n}$ is called a descent direction for Φ at (x,s) when

$$\Phi'(x,s;d_x,d_s) < 0.$$

Assume now that the vector d_x has the form $d_x = y - x$, where y is a point belonging to the domain of f_1; then, from [34, Theorem 23.1] we have

$$\Phi'(x,s;y-x,d_s) \le f_1(y) - f_1(x) + \langle \nabla f_0(x) + x - s, d_x\rangle + \langle s - x, d_s\rangle. \tag{4.49}$$

The above inequality holds independently on the form of d_s.

4.3.1. Descent direction for the merit function. Assume that $(x^{(k)}, s^{(k)})$ is a given point in $\mathrm{dom}(f_1) \times \mathbb{R}^n$, while α_k, β_k are two given parameters. Let the function $h^{(k)}(y;x,s)$ be defined as in (4.26). Given a tolerance parameter $\tau > 0$, according to (4.8)–(4.10), we denote by $\tilde{y}^{(k)}$ any point in $\mathrm{dom}(f_1)$ satisfying

$$0 \in \partial_{\epsilon_k} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}), \quad \text{with } \epsilon_k = -\frac{\tau}{2} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}). \tag{4.50}$$

For a given $\gamma_k \ge 0$, consider

$$d^{(k)} = \begin{pmatrix} d_x^{(k)} \\ d_s^{(k)} \end{pmatrix}, \tag{4.51}$$

where

$$d_x^{(k)} = \tilde{y}^{(k)} - x^{(k)} \tag{4.52}$$

$$d_s^{(k)} = \left(1 + \frac{\beta_k}{\alpha_k}\right)(\tilde{y}^{(k)} - x^{(k)}) + \gamma_k(x^{(k)} - s^{(k)}). \tag{4.53}$$

In the following lemma we show that $d^{(k)}$ is a descent direction for Φ at $(x^{(k)}, s^{(k)})$.

LEMMA 4.12. *Let $d^{(k)} \in \mathbb{R}^{2n}$ be defined according to (4.52)-(4.53), where $0 < \alpha_k \leq \alpha_{max}$, $\beta_k \geq 0$, $\gamma_k \geq \gamma_{min}$, for some given positive real constants $\alpha_{max}, \gamma_{min} > 0$. Define $\Delta_k \in \mathbb{R}_{\leq 0}$ as*

$$\Delta_k = h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) - \gamma_k \|x^{(k)} - s^{(k)}\|^2. \tag{4.54}$$

Then, we have

$$\Phi'(x^{(k)}, s^{(k)}; d_x^{(k)}, d_s^{(k)}) \leq \Delta_k \tag{4.55}$$

$$\leq -a\|\tilde{y}^{(k)} - x^{(k)}\|^2 - \gamma_{min}\|x^{(k)} - s^{(k)}\|^2 \tag{4.56}$$

where $a > 0$ is a positive constant. Therefore, $\Phi'(x^{(k)}, s^{(k)}; d_x^{(k)}, d_s^{(k)}) < 0$ whenever $\tilde{y}^{(k)} \neq x^{(k)}$ or $x^{(k)} \neq s^{(k)}$.

Proof. We first observe that (4.49) with $x = x^{(k)}$, $s = s^{(k)}$, $y = \tilde{y}^{(k)}$, $d_x = d_x^{(k)}$, $d_s = d_s^{(k)}$, gives:

$$\begin{aligned}
&\Phi'(x^{(k)}, s^{(k)}; d_x^{(k)}, d_s^{(k)}) \\
&\leq f_1(\tilde{y}^{(k)}) - f_1(x^{(k)}) + \langle \nabla f_0(x^{(k)}) + x^{(k)} - s^{(k)}, d_x^{(k)} \rangle + \langle s^{(k)} - x^{(k)}, d_s^{(k)} \rangle \qquad (4.57)\\
&= f_1(\tilde{y}^{(k)}) - f_1(x^{(k)}) + \langle \nabla f_0(x^{(k)}), \tilde{y}^{(k)} - x^{(k)} \rangle + \langle x^{(k)} - s^{(k)}, y^{(k)} - x^{(k)} \rangle + \\
&\quad + \left(1 + \frac{\beta_k}{\alpha_k}\right) \langle s^{(k)} - x^{(k)}, \tilde{y}^{(k)} - x^{(k)} \rangle - \gamma_k \|x^{(k)} - s^{(k)}\|^2 \\
&= f_1(\tilde{y}^{(k)}) - f_1(x^{(k)}) + \langle \nabla f_0(x^{(k)}) - \frac{\beta_k}{\alpha_k}(x^{(k)} - s^{(k)}), \tilde{y}^{(k)} - x^{(k)} \rangle - \gamma_k \|x^{(k)} - s^{(k)}\|^2 \\
&\leq h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) - \gamma_k \|x^{(k)} - s^{(k)}\|^2 \qquad (4.58)\\
&\leq -\frac{\theta}{2\alpha_k}\|\tilde{y}^{(k)} - x^{(k)}\|^2 - \gamma_k \|x^{(k)} - s^{(k)}\|^2
\end{aligned}$$

where the last inequality follows from (4.13). Then, the thesis follows from $\alpha_k \leq \alpha_{max}$ with $a = \theta/2\alpha_{max}$. □

A useful consequence of the previous lemma is stated in the following corollary.

COROLLARY 4.13. *Assume that $\alpha_k \in [\alpha_{min}, \alpha_{max}]$, $\beta_k \in [0, \beta_{max}]$, $\gamma_k \in [0, \gamma_{max}]$, with $0 < \alpha_{min} \leq \alpha_{max}$, $\beta_{max} \geq 0$, $\gamma_{max} > 0$. Then, there exists a positive constant C such that*

$$\Delta_k \leq -C\|d^{(k)}\|^2. \tag{4.59}$$

Proof. Setting $\delta_k = 1 + \frac{\beta_k}{\alpha_k}$, the bounds on the parameters imply that $1 \leq \delta_k \leq \bar{\delta}$, with $\bar{\delta} = 1 + \frac{\beta_{max}}{\alpha_{min}}$. By definition of $d^{(k)}$ in (4.51) we have

$$\begin{aligned}
\|d^{(k)}\|^2 &= \|d_x^{(k)}\|^2 + \|d_s^{(k)}\|^2 \\
&= \|\tilde{y}^{(k)} - x^{(k)}\|^2 + \|\delta_k(\tilde{y}^{(k)} - x^{(k)}) + \gamma_k(x^{(k)} - s^{(k)})\|^2 \\
&= (1 + \delta_k^2)\|\tilde{y}^{(k)} - x^{(k)}\|^2 + \gamma_k^2\|x^{(k)} - s^{(k)}\|^2 + 2\delta_k\gamma_k\langle \tilde{y}^{(k)} - x^{(k)}, x^{(k)} - s^{(k)} \rangle \\
&\leq (1 + \delta_k^2 + \delta_k\gamma_k)\|\tilde{y}^{(k)} - x^{(k)}\|^2 + (\gamma_k^2 + \delta_k\gamma_k)\|x^{(k)} - s^{(k)}\|^2 \\
&\leq (1 + \bar{\delta}^2 + \bar{\delta}\gamma_{max})\|\tilde{y}^{(k)} - x^{(k)}\|^2 + \gamma_k(\gamma_{max} + \bar{\delta})\|x^{(k)} - s^{(k)}\|^2 \\
&\leq \frac{1}{C}(a\|\tilde{y}^{(k)} - x^{(k)}\|^2 + \gamma_k\|x^{(k)} - s^{(k)}\|^2)
\end{aligned}$$

where $C = 1/\max\{(1 + \bar{\delta}^2 + \bar{\delta}\gamma_{max})/a, \gamma_{max} + \bar{\delta}\}$. Multiplying both sides of the last inequality above by C and combining with (4.56) gives (4.59). □

Line–Search Armijo line–search

INPUT: $(x^{(k)}, s^{(k)}) \in \mathbb{R}^n \times \mathbb{R}^n$, $d^{(k)}$, Δ_k as in (4.54), $\sigma, \delta \in (0,1)$
Set $\lambda^+ = 1$.
WHILE $\Phi(x^{(k)} + \lambda^+ d_x^{(k)}, s^{(k)} + \lambda^+ d_s^{(k)}) > \Phi(x^{(k)}, s^{(k)}) + \sigma\lambda^+\Delta_k$
Set $\lambda^+ = \delta\lambda^+$
Set $\lambda_k^+ = \lambda^+$
END
OUTPUT: λ_k^+.

4.3.2. Outline of the algorithm. In this section we present a new forward–backward–type method, which combines the inertial approach with the line–search procedure at the basis of VMILA [12, 16]. The convergence guarantees of the new method are based on a line–search along the descent direction defined in the previous section, ensuring the sufficient decrease of the merit function $\Phi(x, s)$. The line–search consists in a backtracking procedure with a generalized Armijo inequality as stopping rule. In particular, if the vector $d^{(k)}$ in (4.51) is a descent direction for Φ at $(x^{(k)}, s^{(k)})$, the line–search algorithm computes a positive parameter λ_k^+ satisfying

$$\Phi(x^{(k)} + \lambda_k^+ d_x^{(k)}, s^{(k)} + \lambda_k^+ d_s^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + \sigma\lambda_k^+\Delta_k, \quad \sigma \in (0,1). \tag{4.60}$$

The well posedness of the line–search procedure and its main properties are summarized in the following lemma.

LEMMA 4.14. *Let the assumptions of Lemma 4.12 be satisfied. Then, the Armijo backtracking line–search algorithm terminates in a finite number of steps, and there exists $\lambda_{min} > 0$ such that the parameter λ_k^+ computed with the line–search algorithm satisfies*

$$\lambda_k^+ \geq \lambda_{min}. \tag{4.61}$$

Proof. Since Φ_0 has M-Lipschitz continuous gradient, with $M = L + 2$ (see (4.47)), we can apply the Descent Lemma obtaining

$$\begin{aligned} \Phi_0(x^{(k)} + \lambda d_x^{(k)}, s^{(k)} + \lambda d_s^{(k)}) &\leq \Phi_0(x^{(k)}, s^{(k)}) + \lambda\langle\nabla\Phi_0(x^{(k)}, s^{(k)}), d^{(k)}\rangle + \frac{M}{2}\lambda^2\|d^{(k)}\|^2 \\ &\leq \Phi_0(x^{(k)}, s^{(k)}) + \lambda\langle\nabla\Phi_0(x^{(k)}, s^{(k)}), d^{(k)}\rangle - \frac{M}{2C}\lambda^2\Delta_k, \end{aligned} \tag{4.62}$$

where the second inequality follows from (4.59). From the Jensen's inequality applied to the convex function f_1 we also obtain

$$\begin{aligned} \Phi_1(x^{(k)} + \lambda d_x^{(k)}, s^{(k)} + \lambda d_s^{(k)}) &= f_1(x^{(k)} + \lambda d_x^{(k)}) = f_1(\lambda\tilde{y}^{(k)} + (1-\lambda)x^{(k)}) \\ &\leq (1-\lambda)f_1(x^{(k)}) + \lambda f_1(\tilde{y}^{(k)}). \end{aligned} \tag{4.63}$$

Summing (4.62) with (4.63) gives

$$\begin{aligned} &\Phi(x^{(k)} + \lambda d_x^{(k)}, s^{(k)} + \lambda d_s^{(k)}) \leq \\ &\quad \leq \Phi(x^{(k)}, s^{(k)}) + \lambda\left(f_1(\tilde{y}^{(k)}) - f_1(x^{(k)}) + \langle\nabla\Phi_0(x^{(k)}, s^{(k)}), d^{(k)}\rangle\right) - \frac{M}{2C}\lambda^2\Delta_k \\ &\quad \leq \Phi(x^{(k)}, s^{(k)}) + \lambda\Delta_k - \frac{M}{2C}\lambda^2\Delta_k, \end{aligned}$$

where the last inequality follows from (4.57)–(4.58). The above relation implies

$$\Phi(x^{(k)} + \lambda d_x^{(k)}, s^{(k)} + \lambda d_s^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + \lambda(1 - \rho\lambda)\Delta_k, \quad \forall\lambda \in [0,1]$$

with $\rho = \frac{M}{2C}$. Moreover, comparing the above inequality with the Armijo condition (4.60) shows that the last one is surely fulfilled when λ_k^+ satisfies $1 - \rho\lambda_k^+ \geq \sigma$, that is when $\lambda_k^+ \leq (1-\sigma)/\rho$. Since λ_k^+ in the backtracking procedure is obtained starting from 1 and by successive reductions

iPila : inertial Proximal inexact line–search algorithm

INPUT: $(x^{(0)}, s^{(0)}) \in \mathrm{dom}(f_1) \times \mathbb{R}^n$, $\sigma \in (0,1)$, $0 < \alpha_{min} \leq \alpha_{max}$, $\beta_{max} > 0$, $0 < \gamma_{min} \leq \gamma_{max}$, $\tau \geq 0$.

FOR $k = 0, 1, \ldots$

STEP 1. Choose $\alpha_k \in [\alpha_{min}, \alpha_{max}]$, $\beta_k \in [0, \beta_{max}]$

STEP 2. Compute $\tilde{y}^{(k)}$ such that

$$0 \in \partial_{\epsilon_k} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}), \quad \text{with } \epsilon_k = -\frac{\tau}{2} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}).$$

STEP 3. Choose $\gamma_k \in [\gamma_{min}, \gamma_{max}]$.

STEP 4. Compute $\Delta_k = h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) - \gamma_k \|x^{(k)} - s^{(k)}\|^2$.

STEP 5. Compute the search direction

$$d_x^{(k)} = \tilde{y}^{(k)} - x^{(k)}$$
$$d_s^{(k)} = \left(1 + \frac{\beta_k}{\alpha_k}\right)(\tilde{y}^{(k)} - x^{(k)}) + \gamma_k (x^{(k)} - s^{(k)})$$

STEP 6. Compute $\lambda_k \in (0, 1]$ such that

$$\Phi(x^{(k)} + \lambda_k d_x^{(k)}, s^{(k)} + \lambda_k d_s^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + \sigma \lambda_k \Delta_k$$

with the line–search backtracking algorithm.

STEP 7. Define the new point as

$$(x^{(k+1)}, s^{(k+1)}) = \begin{cases} (\tilde{y}^{(k)}, x^{(k)}) & \text{if } \Phi(\tilde{y}^{(k)}, x^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + \sigma \lambda_k \Delta_k \\ (x^{(k)} + \lambda_k d_x^{(k)}, s^{(k)} + \lambda_k d_s^{(k)}) & \text{otherwise} \end{cases}$$

END

of a factor $\delta < 1$, we have $\lambda_k^+ \geq \delta^M$, where M is the smallest nonnegative integer such that $\delta^M \leq (1 - \sigma)/\rho$. Therefore, (4.61) is satisfied with $\lambda_{min} = \delta^M$. □

The descent direction and the backtracking procedure described above are at the basis of the new algorithm, named iPila (inertial Proximal inexact line–search algorithm), which formally consists in a descent method for the merit function $\Phi(x, s)$. In particular, it generates a sequence of iterates $\{(x^{(k)}, s^{(k)})\}_{k \in \mathbb{N}}$ and a sequence of steplength parameters $\{\lambda_k\}_{k \in \mathbb{N}}$ fulfilling the following decrease condition

$$\Phi(x^{(k+1)}, s^{(k+1)}) \leq \Phi(x^{(k)}, s^{(k)}) + \sigma \lambda_k \Delta_k \quad \text{and} \quad \Phi(x^{(k+1)}, s^{(k+1)}) \leq \Phi(\tilde{y}^{(k)}, x^{(k)}).$$

The connection with the inertial methods is in fact that, when $(x^{(k+1)}, s^{(k+1)}) = (\tilde{y}^{(k)}, x^{(k)})$ is selected at STEP 7, the following iteration will consist of an actual inertial step. In practice, the condition at STEP 7 can be considered as an alternative acceptance rule for the inexact inertial proximal gradient point, having a similar role than the condition at STEP 6 of i^{2}Piano (see also (4.28)). The main difference is that here the acceptance condition is based on the Armijo inequality, while the one in i^{2}Piano is based on the Descent Lemma.

Notice also that the inexact evaluation of the proximity operator in iPila is required only once per iteration, unlike in i^{2}Piano, where it is needed at each step of the loop for selecting the parameter L_k, until inequality (4.28) is satisfied.

The Armijo condition results also in a larger freedom of choosing the parameters α_k, β_k, which here satisfy very minimal conditions. A possible strategy to choose these parameters preserving both the theoretical prescriptions and the benefits deriving from the presence of an inertial step is described in Section 5.

4.3.3. Convergence analysis. In this section we first show that conditions *[H1]–[H3]* are satisfied for iPila.

PROPOSITION 4.15. *Let $\{(x^{(k)}, s^{(k)})\}_{k\in\mathbb{N}}$ be the sequence generated by iPila. Then, condition [H1] holds with $d_k = \sqrt{-\Delta_k}$, $a_k = \sigma\lambda_{min}$. Moreover, under Assumption [A4], we also have*

$$0 = \lim_{k\to\infty} \|x^{(k)} - s^{(k)}\| = \lim_{k\to\infty} h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) = \lim_{k\to\infty} \|\tilde{y}^{(k)} - x^{(k)}\|. \tag{4.64}$$

Proof. From the updating rule at STEP 7 and from Lemma 4.14, we have

$$\Phi(x^{(k+1)}, s^{(k+1)}) \le \Phi(x^{(k)}, s^{(k)}) + \sigma\lambda_k\Delta_k \le \Phi(x^{(k)}, s^{(k)}) + \sigma\lambda_{min}\Delta_k.$$

Then, condition *[H1]* is satisfied with $d_k^2 = -\Delta_k$, $a_k = \sigma\lambda_{min}$. Since from Assumption *[A4]* f is bounded from below, Φ is bounded from below as well. Therefore, *[H1]* implies $-\sum_{k=0}^{\infty}\Delta_k < \infty$ which, in turn, yields $\lim_{k\to\infty}\Delta_k = 0$. Recalling (4.56), this implies (4.64). □

In the following we describe the setup for proving *[H2]*, with the second auxiliary function defined as in (4.39).

PROPOSITION 4.16. *Let $\{(x^{(k)}, s^{(k)})\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm iPila with $\gamma_{min} > 0$ and let $\mathcal{F}$ be defined as in (4.39). Then, there exist $\{\rho^{(k)}\}_{k\in\mathbb{N}}, \{r_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$, with $\lim_{k\to\infty} r_k = 0$, such that [H2] holds with $u^{(k)} = \hat{y}^{(k)}$, where $\hat{y}^{(k)}$ is the exact minimizer of $h^{(k)}(y; x^{(k)}, s^{(k)})$, i.e.,*

$$\hat{y}^{(k)} = \operatorname*{argmin}_{y\in\mathbb{R}^n} h^{(k)}(y; x^{(k)}, s^{(k)}).$$

Proof. From STEP 7, we have

$$\begin{aligned}\Phi(x^{(k+1)}, s^{(k+1)}) &\le \Phi(\tilde{y}^{(k)}, x^{(k)}) = f(\tilde{y}^{(k)}) + \frac{1}{2}\|\tilde{y}^{(k)} - x^{(k)}\|^2 \\ &\le f(\hat{y}^{(k)}) - \left(c + \frac{\alpha_{max}}{\theta}\right) h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) + d\|x^{(k)} - s^{(k)}\|^2,\end{aligned}$$

where the last inequality follows from (4.16) and (4.13). Setting

$$\rho^{(k)} = \sqrt{2}\left(-\left(c + \frac{\alpha_{max}}{\theta}\right) h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) + d\|x^{(k)} - s^{(k)}\|^2\right)^{\frac{1}{2}}, \tag{4.65}$$

we obtain $\Phi(x^{(k+1)}, s^{(k+1)}) \le \mathcal{F}(\hat{y}^{(k)}, \rho^{(k)})$, which represents the left-most inequality in *[H2]*, with $u^{(k)} = \hat{y}^{(k)}$. On the other hand, from inequality (4.17) we obtain

$$\mathcal{F}(\hat{y}^{(k)}, \rho^{(k)}) = f(\hat{y}^{(k)}) + \frac{1}{2}(\rho^{(k)})^2 \le f(x^{(k)}) - \bar{c}h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) + \bar{d}\|x^{(k)} - s^{(k)}\|^2 + \frac{1}{2}(\rho^{(k)})^2.$$

Setting $r_k = (\rho^{(k)})^2/2 - \bar{c}h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)}) + (\bar{d} - \frac{1}{2})\|x^{(k)} - s^{(k)}\|^2$, we can write

$$\mathcal{F}(\hat{y}^{(k)}, \rho^{(k)}) \le f(x^{(k)}) + \frac{1}{2}\|x^{(k)} - s^{(k)}\|^2 + r_k = \Phi(x^{(k)}, s^{(k)}) + r_k.$$

From (4.64) we have that $\lim_{k\to\infty} r_k = 0$ and this proves *[H2]*. □

PROPOSITION 4.17. *Let $\{(x^{(k)}, s^{(k)})\}_{k\in\mathbb{N}}$ be the sequence generated by iPila with $\gamma_{min} > 0$. Then, there exists a positive constant b such that [H3] is satisfied with $I = \{1\}$, $\theta_1 = 1$, $\zeta_k = 0$.*

Proof. From (4.15) we know that there exists a subgradient $\hat{v}^{(k)} \in \partial f(\hat{y}^{(k)})$ such that

$$\|\hat{v}^{(k)}\| \le q\sqrt{-h^{(k)}(\tilde{y}^{(k)}; x^{(k)}, s^{(k)})} + q\|x^{(k)} - s^{(k)}\| \tag{4.66}$$

and, reasoning as in the proof of Proposition 4.8, it follows that

$$\|\partial\mathcal{F}(\hat{y}^{(k)}, \rho^{(k)})\|_- \le \left\|\begin{pmatrix}\hat{v}^{(k)} \\ \rho^{(k)}\end{pmatrix}\right\| \le \|\hat{v}^{(k)}\| + |\rho^{(k)}|. \tag{4.67}$$

Let us analyze the two terms at the right-hand side of the inequality above, showing that both can be bounded from above with a multiple of $\sqrt{-\Delta_k}$. From (4.66) we obtain

$$\|\hat{v}^{(k)}\| \leq q\sqrt{-h^{(k)}(\tilde{y}^{(k)};x^{(k)},s^{(k)})+\gamma_{min}\|x^{(k)}-s^{(k)}\|^2} \\ +\frac{q}{\sqrt{\gamma_{min}}}\sqrt{\gamma_{min}\|x^{(k)}-s^{(k)}\|^2-h^{(k)}(\tilde{y}^{(k)};x^{(k)},s^{(k)})}.$$

which, setting $A = q\left(1+1/\sqrt{\gamma_{min}}\right)$ and using (4.56), yields

$$\|\hat{v}^{(k)}\| \leq A\sqrt{-h^{(k)}(\tilde{y}^{(k)};x^{(k)},s^{(k)})+\gamma_{min}\|x^{(k)}-s^{(k)}\|^2} \leq A\sqrt{-\Delta_k}. \tag{4.68}$$

On the other hand, from definition (4.65)

$$\rho^{(k)} \leq B\sqrt{-h^{(k)}(\tilde{y}^{(k)};x^{(k)},s^{(k)})+\gamma_k\|x^{(k)}-s^{(k)}\|^2} = B\sqrt{-\Delta_k}$$

where $B = \sqrt{2}\max\{c+\alpha_{max}/\theta, d/\gamma_{min}\}^{\frac{1}{2}}$. Therefore, combining the last inequality above with (4.68) and (4.67), yields $\|\partial\mathcal{F}(\hat{y}^{(k)},\rho^{(k)})\|_- \leq (A+B)\sqrt{-\Delta_k}$ which proves that *[H3]* is satisfied with $I=\{1\}$, $\theta_1 = 1$, $\zeta_k = 0$, $b = A+B$, $b_k = 1$. □

We are now ready for presenting the main convergence result for Algorithm iPila.

THEOREM 4.18. *Suppose that $\mathcal{F}$ is a KL function. Moreover, assume that the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ generated by iPila is bounded. Then, $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to a stationary point of f.*

Proof. By Proposition 4.15, 4.16, and 4.17, we know that conditions *[H1]-[H2]-[H3]* hold for iPila. From (4.11) we have $\lim_{k\to\infty}\|u^{(k)}-x^{(k)}\| = \lim_{k\to\infty}\|\hat{y}^{(k)}-x^{(k)}\| = 0$. Hence we can apply Proposition 4.8 and conclude that *[H4]* holds. Moreover, condition *[H5]* holds as a consequence of Lemma 4.12, since $d_k = \sqrt{-\Delta_k}$ and $\|\tilde{y}^{(k)}-x^{(k)}\| \geq \|x^{(k+1)}-x^{(k)}\|/\lambda_k \geq \|x^{(k+1)}-x^{(k)}\|$ (see STEP 7). Finally, condition *[H6]* is trivially satisfied, since both sequences $\{a_k\}_{k\in\mathbb{N}}$ and $\{b_k\}_{k\in\mathbb{N}}$ are constant. Then Theorem 3.3 applies and guarantees that the sequence $\{(x^{(k)},\rho^{(k)})\}_{k\in\mathbb{N}}$ converges to a stationary point (x^*,ρ^*) of $\mathcal{F}$. Note that, since $\mathcal{F}$ is the sum of separable functions, its subdifferential can be written as $\partial\mathcal{F}(x,\rho) = \partial f(x)\times\{\rho\}$. Then, (x^*,ρ^*) is stationary for $\mathcal{F}$ if and only if $\rho^* = 0$ and $0\in\partial f(x^*)$. Hence x^* is a stationary point for f and $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to it. □

We refer the reader to Remark 4.10-4.11 for conditions on f guaranteeing that $\mathcal{F}$ is a KL function and that the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ is bounded.

5. Numerical illustration: image denoising and deblurring in presence of impulse noise. In image restoration problems the goal is to recover a good quality image from a noisy blurred one. Following the variational approach, the clean image is obtained by solving an optimization problem with the structure (4.1), where the objective function includes a measure of the data fidelity and a regularization/penalization term, incorporating all a priori information on the desired solution. The data discrepancy is usually selected according to the noise statistics: in particular, when the data are affected by impulse noise, the variational model is defined as

$$\min_{x\in\mathbb{R}^n_{\geq 0}} \|Hx-g\|_1 + \mathcal{R}(x),$$

where $g\in\mathbb{R}^n$ is the noisy blurred data, $H\in\mathbb{R}^{n\times n}$ represents the blurring operator, $\mathcal{R}:\mathbb{R}^n\to\mathbb{R}$ is the regularization term, and $\mathbb{R}^n_{\geq 0} = \{x\in\mathbb{R}^n : x_i\geq 0, i=1,\ldots,n\}$ is the nonnegative orthant. In the recent literature, several nonconvex regularization functionals have been proposed to overcome the known drawbacks of classical approaches, for example those based on the Total Variation function. In this paper we consider as regularization function the one proposed in [20, 21]:

$$\mathcal{R}(x) = \rho\sum_{\ell=1}^{q}\sum_{i=1}^{n}\log(1+(K_\ell x)_i^2),$$

where the matrices $K_\ell \in \mathbb{R}^{n\times n}$ correspond to the convolution with a given filter k_ℓ, while ρ, θ_ℓ are positive parameters. In particular, the set of 48 filters k_ℓ of size 7×7 and corresponding coefficients θ_ℓ are computed with a supervised machine learning technique. It is possible to prove that $\mathcal{R}(x)$ has Lipschitz-continuous gradient, using the same arguments as in [16]. Then, setting $f_0(x) = \mathcal{R}(x)$ complies with assumptions *[A2]-[A3]*. The nonnegativity constraint, expressed by means of the indicator function of the nonnegative orthant $\iota_{\geq 0}(x)$, can be included in the convex, nonsmooth term of the objective function, i.e. $f_1(x) = \|Hx - g\|_1 + \iota_{\geq 0}(x)$. The regularization parameter ρ has been manually tuned in order to have a good quality restoration. Its value has been set equal to 0.08 for all the runs. Note that f_1 does not have a closed-form proximal operator, therefore it is necessary to employ implementable inexactness criteria as the one proposed in Section 4.1.
The proposed algorithms have been implemented in Matlab on a laptop equipped with a 2.5 GHz Intel Core i7-6500 processor and 16GB of RAM; the Matlab code is available online at [13].
The parameters of Algorithm i^{2}Piano have been set as $\delta = 0.5$, $\gamma = 0.2$, $\eta = 1.5$, $\omega = 0.95$. The estimate of the Lipschitz constant L_k is updated in a nondecreasing way. In particular, the initial value L_0 is set as an input parameter: then, at Step 1 of each iteration, the first tentative value is set as $L_k = L_{k-1}$. This value is possibly increased until inequality (4.28) is met. Actually, more sophisticated updating rules for this parameter could be adopted; however the objective function of the considered image restoration problem is very costly to evaluate, therefore a more conservative parameters selection rule has shown to be more convenient. As for Algorithm iPila, the parameters settings aim to mimic that of the inertial method i^{2}Piano. Indeed, introducing the additional parameters $\delta, \gamma, L_k, b_k > 0$, with $b_k = \frac{L_k + 2\delta}{L_k + 2\gamma}$, we set $\alpha_k, \beta_k, \gamma_k$ as follows

$$\beta_k = \frac{b_k - 1}{b_k - \frac{1}{2}}, \qquad \alpha_k = 2\frac{1-\beta_k}{L_k + 2\gamma}, \qquad \gamma_k = \gamma.$$

This choice is motivated by the following arguments. If L_k is a good local approximation of the Lipschitz constant satisfying condition

$$f_0(\tilde{y}^{(k)}) \leq f_0(x^{(k)}) + \langle \nabla f_0(x^{(k)}), \tilde{y}^{(k)} - x^{(k)} \rangle + \frac{L_k}{2}\|\tilde{y}^{(k)} - x^{(k)}\|^2, \tag{5.1}$$

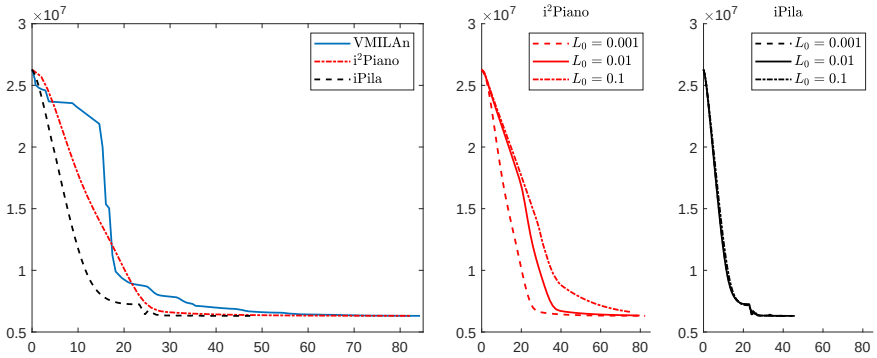
then reasoning as in the proof of Proposition 4.5, and choosing $\delta = 0.5$ and $\gamma_k = \gamma$, we obtain

$$\Phi(\tilde{y}^{(k)}, x^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + \Delta_k + \frac{1}{2\alpha_k}\|\tilde{y}^{(k)} - x^{(k)}\|^2.$$

Hence, if L_k satisfies (5.1), the point $(\tilde{y}^{(k)}, x^{(k)})$ will be likely accepted at Step 7, as also confirmed by the numerical experience. This reasoning suggests to implement algorithm iPila as follows. We check the condition $\Phi(\tilde{y}^{(k)}, x^{(k)}) \leq \Phi(x^{(k)}, s^{(k)}) + \sigma\Delta_k$ right after Step 4: if the condition holds, then the steps from 4 to 7 are skipped in order to avoid unnecessary computations, and the next point is directly defined as $(x^{(k+1)}, s^{(k+1)}) = (\tilde{y}^{(k)}, x^{(k)})$; otherwise, the value L_k is increased by a factor $\eta = 1.5$, and the line–search in steps 5–7 is performed in order to compute the next point. By possibly increasing L_k, we aim at improving the chances that $(\tilde{y}^{(k+1)}, x^{(k+1)})$ is accepted at the next iteration, thus reducing the computational time due to the line–search reductions steps. The parameter in the Armijo condition, is set to $\sigma = 10^{-4}$.
For both i^{2}Piano and iPila, the inexact proximal point $\tilde{y}^{(k)}$ is computed by approximately solving the dual of problem $\min_{y\in\mathbb{R}^n} h^{(k)}(y; x^{(k)}, s^{(k)})$, which is a quadratic problem with simple constraints, with FISTA (more details can be found in [16] and references therein). The accuracy of the approximation is controlled by the parameter τ: in our experiments we set $\tau = 10^6$, which corresponds to a good balancing of the computational complexity among inner and outer iterations. An extensive performance assessment of the algorithms with respect to this and other parameters is out of the scope of this paper, and it will be subject of future research.
The deblurring test problem has been obtained by first artificially blurring a good quality image, then simulating impulse noise on the 15% of the pixels with `imnoise`. The clean and the noisy image are in Figure 1 (a) and (b). Assuming reflective boundary conditions, matrix-vector multiplications involving H and H^T can be implemented efficiently with the DCT transform. In particular, each inner (dual) iteration requires the computation of two matrix vector product of this kind. Indeed, in our experiments, only one or two inner iterations per outer iteration are,

(a) (b) (c)

FIG. 1. *Image deblurring test problem: (a) Original image (*512×512 *pixels); (b) Noisy image, PSNR = 13.19; (c) Restored image, PSNR = 23.72*



(a) (b)

FIG. 2. *Image deblurring test problem: decrease of the objective function (a) with respect to the computational time and (b) with different choices of* L_0.

in general, needed to satisfy the inner stopping criterion. We compare our proposed algorithms to the variable metric line–search based method denominated VMILAn [12, 16] with its standard parameters settings. The objective function decrease for all three algorithms i^{2}Piano, iPila and VMILAn is reported in Figure 2. Panel (a) reports the values $f(x^{(k)})$ with respect to the computational time, where the initial estimate of the Lipschitz constant has been set equal to 0.001 for both i^{2}Piano and iPila. The decrease of the objective function during the iterates of i^{2}Piano and iPila with different choices of the parameter L_0 is also presented in panel (b), where the horizontal axis still refers to the computational time. The picture shows that iPila is quite insensitive to the choice of L_0. In general, the numerical results show that i^{2}Piano and iPila are able to solve challenging problems and are competitive with state-of-the-art methods. More effective rules for selecting their parameters will be subject of future work.

Acknowledgments. Silvia Bonettini, Marco Prato and Simone Rebegoldi are members of the INdAM research group GNCS. Peter Ochs acknowledges funding by the German Research Foundation (DFG Grant OC 150/1-1).

REFERENCES

[1] P. A. Absil, R. Mahony, and B. Andrews, *Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim., 16 (2005), pp. 531–547.

[2] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129.

[3] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, *Structured sparsity through convex optimization*, Stat. Sci., 27 (2012), pp. 450–468.

[4] M. Bertero, P. Boccacci, and V. Ruggiero, *Inverse Imaging with Poisson Data*, IOP Publishing, Bristol, 2018.

[5] D. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, 1999.
[6] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, *Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity*, Trans. Am. Math. Soc., 362 (2010), pp. 3319–3363.
[7] J. Bolte, A. Daniilidis, and S. M., *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 10 (2007), pp. 556–572.
[8] J. Bolte, A. Danilidis, and A. Lewis, *The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2007), pp. 1205–1223.
[9] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.
[10] J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd, *First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems*, SIAM J. Optim., 28 (2018), pp. 2131–2151.
[11] 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.
[12] S. Bonettini, I. Loris, F. Porta, M. Prato, and S. Rebegoldi, *On the convergence of a linesearch based proximal-gradient method for nonconvex optimization*, Inverse Probl., 33 (2017), p. 055005.
[13] S. Bonettini, P. Ochs, M. Prato, and S. Rebegoldi, *inertial inexact Proximal algorithm for nonconvex optimization (i2Piano) and inertial Proximal inexact linesearch algorithm (iPila) software*. http://www.oasis.unimore.it/site/home/software.html, 2021.
[14] S. Bonettini, M. Prato, and S. Rebegoldi, *A block coordinate variable metric linesearch based proximal gradient method*, Comput. Optim. Appl., 71 (2018), pp. 5–52.
[15] S. Bonettini, M. Prato, and S. Rebegoldi, *Convergence of inexact forward-backward algorithms using the forward-backward envelope*, SIAM J. Optim., 30 (2020), pp. 3069–3097.
[16] S. Bonettini, M. Prato, and S. Rebegoldi, *New convergence results for the inexact variable metric forward-backward method*, Appl. Math. Comput., 392 (2021), p. 125719.
[17] S. Bonettini, S. Rebegoldi, and V. Ruggiero, *Inertial variable metric techniques for the inexact forward-backward algorithm*, SIAM J. Sci. Comput., 40 (2018), pp. A3180–A3210.
[18] L. Bottou, F. E. Curtis, and J. Nocedal, *Optimization methods for large-scale machine learning*, SIAM Rev., 60 (2018), pp. 223–311.
[19] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319.
[20] Y. Chen, T. Pock, R. Ranftl, and H. Bischof, *Revisiting loss-specific training of filter-based MRFs for image restoration*, in Pattern Recognition, J. Weickert, M. Hein, and B. Schiele, eds., Berlin, Heidelberg, 2013, Springer Berlin Heidelberg, pp. 271–281.
[21] Y. Chen, R. Ranftl, and T. Pock, *Insights into analysis operator learning: From patch-based sparse models to higher order mrfs*, IEEE Trans. Image Process., 23 (2014), pp. 1060–1072.
[22] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*, J. Optim. Theory Appl., 162 (2014), pp. 107–132.
[23] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, *A block coordinate variable metric forward-backward algorithm*, Journal of Global Optimization, 66 (2016), pp. 457–485.
[24] P. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
[25] 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.
[26] P. Frankel, G. Garrigos, and J. Peypouquet, *Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates*, J. Opt. Theory Appl., 165 (2015), pp. 874–900.
[27] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier, 48 (1998), pp. 769–783.
[28] G. Li and T. K. Pong, *Douglas–Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems*, Math. Program., 159 (2016), pp. 371–401.
[29] D. Noll, *Convergence of non-smooth descent methods using the kurdyka–Łojasiewicz inequality*, J. Opt. Theory Appl., 160 (2014), pp. 553–572.
[30] P. Ochs, *Unifying abstract inexact convergence theorems and block coordinate variable metric iPiano*, SIAM J. Optim., 29 (2019), pp. 541–570.
[31] 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.
[32] B. Polyak, *Introduction to optimization*, Optimization Software - Inc., Publication Division, New York, 1987.
[33] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys., 4 (1964), pp. 1–17.
[34] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[35] R. T. Rockafellar, R. J.-B. Wets, and M. Wets, *Variational Analysis*, vol. 317 of Grundlehren der Mathematischen Wissenschaften, Springer, Berlin, 1998.
[36] L. Stella, A. Themelis, and P. Patrinos, *Forward-backward quasi-Newton methods for nonsmooth optimization problems.*, Comput. Optim. Appl., 67 (2017), pp. 443–487.
[37] S. Villa, S. Salzo, L. Baldassarre, and A. Verri, *Accelerated and inexact forward-backward algorithms*, SIAM J. Optim., 23 (2013), pp. 1607–1633.
[38] A. Zalinescu, *Convex analysis in general vector spaces*, World Scientific Publishing Co. Inc., 2002.