



A nested primal–dual iterated Tikhonov method for regularized convex optimization

Stefano Aleotti¹ · Silvia Bonettini² · Marco Donatelli¹ · Marco Prato² · Simone Rebegoldi² 

Received: 13 January 2024 / Accepted: 21 September 2024 / Published online: 30 September 2024
© The Author(s) 2024

Abstract

Proximal–gradient methods are widely employed tools in imaging that can be accelerated by adopting variable metrics and/or extrapolation steps. One crucial issue is the inexact computation of the proximal operator, often implemented through a nested primal–dual solver, which represents the main computational bottleneck whenever an increasing accuracy in the computation is required. In this paper, we propose a nested primal–dual method for the efficient solution of regularized convex optimization problems. Our proposed method approximates a variable metric proximal–gradient step with extrapolation by performing a prefixed number of primal–dual iterates, while adjusting the steplength parameter through an appropriate backtracking procedure. Choosing a prefixed number of inner iterations allows the algorithm to keep the computational cost per iteration low. We prove the convergence of the iterates sequence towards a solution of the problem, under a relaxed monotonicity assumption on the scaling matrices and a shrinking condition on the extrapolation parameters. Furthermore, we investigate the numerical performance of our proposed method by equipping it with a scaling matrix inspired by the Iterated Tikhonov method. The numerical results

Silvia Bonettini, Marco Donatelli, Marco Prato and Simone Rebegoldi have contributed equally to this work.

✉ Simone Rebegoldi
simone.rebegoldi@unimore.it

Stefano Aleotti
saleotti@studenti.uninsubria.it

Silvia Bonettini
silvia.bonettini@unimore.it

Marco Donatelli
marco.donatelli@uninsubria.it

Marco Prato
marco.prato@unimore.it

¹ Dipartimento di Scienza e Alta Tecnologia, Università dell’Insubria, Via Valleggio 11, 22100 Como, Italy

² Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/B, 41125 Modena, Italy

show that the combination of such scaling matrices and Nesterov-like extrapolation parameters yields an effective acceleration towards the solution of the problem.

Keywords Primal–dual methods · Iterated Tikhonov · Convex optimization · Image deblurring

1 Introduction

We consider the following problem

$$\operatorname{argmin}_{u \in \mathbb{R}^d} f(u) + h(Wu), \quad (1)$$

where $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex and smooth, $h : \mathbb{R}^{d'} \rightarrow \mathbb{R} \cup \{\infty\}$ is convex and possibly nonsmooth, and $W \in \mathbb{R}^{d' \times d}$ is a linear operator. Problem (1) lies at the core of several inverse problems arising in imaging, such as image deblurring, denoising, inpainting, compressed sensing, super-resolution and others [1–4]. We may think of the term f as a discrepancy measure between the data and the imaging model, e.g. the least squares cost function, whereas we may interpret $h \circ W$ as a penalty term enforcing some regularization property on the problem, e.g. a vectorial norm composed with either the identity or the gradient operator.

First order methods are particularly useful to solve problem (1) with a mild-to-moderate accuracy while keeping the computational cost per iteration low, especially when the size d is large. Among these is the proximal–gradient method [5–7], which alternates a gradient step on f and a proximal evaluation on $h \circ W$ as follows

$$\begin{aligned} u_{n+1} &= \operatorname{prox}_{\alpha h \circ W}(u_n - \alpha \nabla f(u_n)) \\ &= \operatorname{argmin}_{u \in \mathbb{R}^d} \frac{1}{2} \|u - (u_n - \alpha \nabla f(u_n))\|^2 + \alpha h(Wu), \quad \forall n \geq 0, \end{aligned} \quad (2)$$

where $\alpha > 0$ is the steplength parameter along the descent direction $-\nabla f(u_n)$ and $\operatorname{prox}_{\alpha h \circ W}$ is the proximal operator of the nonsmooth term $\alpha h \circ W$. The convergence of method (2) to a solution of (1) is ensured whenever f has a L –Lipschitz continuous gradient and α is smaller than $2/L$, see for instance [6].

The scheme (2) has two practical drawbacks. On the one hand, it might converge slowly to the desired solution whenever the prefixed steplength α is too small, which typically happens when only a coarse (large) estimate of the Lipschitz constant L is available. A possible remedy to this issue consists in accelerating the scheme by either computing the proximal–gradient step with respect to a variable metric that encapsulates some second order information of the differentiable part [8–12], or inserting an extrapolation step that exploits the information relative to the previous iterates [13, 14]. On the other hand, the scheme (2) is based on the implicit assumption that $\operatorname{prox}_{\alpha h \circ W}$ is computable in closed form, which excludes several regularization terms such as Total Variation, overlapping group Lasso or fused Lasso penalties [1, 15]. Alternative splitting approaches that circumvent the explicit evaluation of $\operatorname{prox}_{\alpha h \circ W}$ include

primal–dual methods based on the reformulation of (1) as a convex–concave saddle point problem [16–18], proximal–gradient methods with inexact proximal evaluations computed through a primal–dual inner solver [8, 19, 20], and the alternating direction method of multipliers (ADMM) and its variants based on the alternating minimization of the augmented Lagrangian [21–23].

In the following, we focus our attention on proximal–gradient methods where acceleration techniques based on variable metrics and extrapolation are combined and inexact proximal evaluations are allowed, namely

$$\begin{cases} \bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}) \\ u_{n+1} \approx \text{prox}_{\alpha_n h \circ W}^{P_n}(\bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n)), \end{cases} \quad \forall n \geq 0, \tag{3}$$

where $\alpha_n > 0$ and $P_n \in \mathbb{R}^{d \times d}$ are, respectively, the steplength parameter and the symmetric positive definite scaling matrix, $\text{prox}_{\alpha_n h \circ W}^{P_n}$ is the proximal operator of $\alpha_n h \circ W$ with respect to the norm induced by P_n , and $\gamma_n \geq 0$ is the extrapolation parameter. The symbol “ \approx ” denotes an approximation of the proximal–gradient point, which is required whenever the proximal operator is not computable in closed form.

The parameters α_n and P_n define the *variable metric* with respect to which the proximal–gradient point u_{n+1} in (3) is computed. While α_n represents the inverse of a local Lipschitz constant of the gradient that is dynamically computed through a backtracking procedure [5], the matrix P_n aims at identifying some second order information of the smooth part of the objective function. Practical choices for P_n include the Hessian matrix or its regularized versions [24, 25], Hessian approximations based on Quasi-Newton strategies [11, 12, 26, 27], or diagonal matrices obtained by the split gradient strategy for nonnegatively constrained problems [28–30]. The extrapolation parameter γ_n is usually computed according to the prefixed sequence formerly proposed for smooth problems by Nesterov [31] and then successfully adapted to non-smooth problems by Beck and Teboulle [5], which guarantees an optimal $\mathcal{O}(1/n^2)$ convergence rate for the function values.

The approximation of the proximal operator in (3) is typically achieved by means of a *nested* iterative solver, which is applied, at each iteration, to the minimization problem associated to the computation of the proximal–gradient point. If the convex conjugate $h^*(v) = \sup_{u \in \mathbb{R}^{d'}} \langle v, u \rangle - h(u)$ has an easy-to-compute proximal operator, then the desired proximal approximation can be computed by means of a *primal–dual* routine that involves only the computation of $\nabla f(u_n)$, prox_{h^*} and the matrix–vector products with W , W^T , P_n^{-1} [20, 29, 32]. In this respect, we can distinguish between two different approaches in the literature. On the one hand, one could approximate the proximal–gradient point with increasing accuracy, which means that the number of inner iterations of the nested solver grows unbounded as the outer iterations proceed [11, 20, 26, 29, 33]. The main drawback of this approach is that the computational cost per iteration may increase in an unsustainable manner, leading to a computational bottleneck in a few iterations. On the other hand, one could compute the approximated proximal–gradient point by means of a prefixed number of inner iterations, while employing an appropriate starting condition for the inner solver; this is advantageous in order to keep the computational cost per iteration low and fixed. Some methods

exploiting this latter approach are available in the literature; see e.g. [13], which considers an inexact version of the popular FISTA that corresponds to (3) with P_n equal to the identity matrix, γ_n selected according to Nesterov's sequence, and u_{n+1} computed through a prefixed number of primal–dual inner iterates with null initialization for the inner solver; [32], where the scheme (3) is presented without extrapolation nor variable metrics and is equipped with a prefixed number of primal–dual inner iterates, which are warm-started with the outcome of the previous outer iteration; [34], where the authors generalize the algorithm in [32] by adding an extrapolation step. Nonetheless, to the best of our knowledge, a method of the form (3) equipped with a prefixed number of inner iterations that include simultaneously variable steplengths, metrics, and extrapolation, has yet to be proposed in the literature.

Contribution. In this paper, we devise and analyse a nested primal–dual method based on the scheme (3) for solving problem (1). At every iteration, our proposed method first performs an extrapolation step on the previous iterates and then computes a prefixed number of primal–dual iterates to approximate a variable metric proximal–gradient step taken from the extrapolated iterate. Furthermore, the steplength parameter is adjusted by means of a backtracking procedure based on a local approximation of the Lipschitz constant of the gradient. From the theoretical viewpoint, we prove the convergence of the iterates sequence towards a minimum point of problem (1), under a relaxed monotonicity assumption on the scaling matrices and a shrinking condition on the extrapolation parameters. Numerically, we focus on the solution of Total-Variation regularized least–squares problems of the form

$$\operatorname{argmin}_{u \in \mathbb{R}^d} \frac{1}{2} \|Au - b\|^2 + TV(u), \quad (4)$$

where $A \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$, and $TV(u)$ denotes the Total Variation function [35]. We investigate the performance of our proposed nested primal–dual algorithm for solving (4) by equipping it with the following scaling matrix

$$P_n = A^T A + \nu_n I_d, \quad \forall n \geq 0, \quad (5)$$

where $I_d \in \mathbb{R}^{d \times d}$ is the identity matrix and $\{\nu_n\}_{n \in \mathbb{N}}$ is either a positive constant or a monotone sequence of positive real numbers. Such a matrix is inspired by works on regularizing preconditioning for image deblurring, drawing inspiration from the Iterated Tikhonov method [36–38]. We show that the combination of such preconditioning matrices and Nesterov-like extrapolation parameters yields an effective acceleration towards the solution of the problem.

Related work. Note that our proposed method generalizes the nested primal–dual method with extrapolation in [34], by allowing for the presence of variable metrics and steplengths. Other inexact variable metric proximal–gradient algorithms with extrapolation that can be cast in the scheme (3) have been proposed in [11, 29]. Our proposed algorithm differs from these two competitors in at least three aspects. (i) We compute the proximal–gradient point with a prefixed number of primal–dual iterates, whereas in [11, 29] the number of inner iterations increases unbounded as the outer iterations

proceed. (ii) The extrapolation parameter in [11, 29] coincides precisely with Nesterov’s sequence, whereas in our approach the extrapolation parameter must satisfy an additional shrinking condition for theoretical purposes. (iii) In the numerical experiments, the scaling matrix P_n in [11, 29] is computed by following either Quasi-Newton or split gradient strategies, whereas our proposed algorithm computes P_n according to an Iterated Tikhonov approach. Regarding the latter aspect, we remark that regularized Hessian matrices (such as the one in (5)) have already been considered in the context of proximal–gradient algorithms, see e.g. [12, 25, 27]. However, to the best of our knowledge, the effect of Hessian regularization combined with extrapolation has yet to be extensively explored in the literature. In our numerical experiments, we investigate the combination of these two techniques within our proposed algorithm for solving the regularized problem (4) in the context of image deblurring.

Paper organization. The paper is organized as follows. In Sect. 2 we recall some basic notions of convex analysis. Section 3 is devoted to presenting our algorithm and analyzing its convergence under suitable hypotheses on the extrapolation parameter and the variable metric. Section 4 concerns the application of our algorithm to Total Variation based image deblurring. Conclusions and future work are discussed in Sect. 5.

2 Preliminaries

Throughout the paper, we denote with $\mathbb{R}^{d \times d}$ the space of $d \times d$ real-valued matrices and $I_d \in \mathbb{R}^{d \times d}$ the $d \times d$ identity matrix. The symbol $\langle \cdot, \cdot \rangle$ denotes the standard inner product on \mathbb{R}^d . If $v \in \mathbb{R}^d$ and $W \in \mathbb{R}^{d \times d}$, $\|v\| = \sqrt{\langle v, v \rangle}$ is the Euclidean norm of v , whereas $\|W\|$ is the spectral norm of W , i.e., the largest singular value of W . We denote with $\mathcal{S}(\mathbb{R}^d)$ the set of all $d \times d$ symmetric matrices, and $\mathcal{S}_+(\mathbb{R}^d)$ the set of all $d \times d$ symmetric positive definite matrices. We recall the definition of the Loewner partial ordering relation, i.e.,

$$\forall P_1, P_2 \in \mathcal{S}(\mathbb{R}^d) \quad P_1 \leq P_2 \iff \langle u, P_1 u \rangle \leq \langle u, P_2 u \rangle, \quad \forall u \in \mathbb{R}^d.$$

Given $P \in \mathcal{S}_+(\mathbb{R}^d)$, the norm induced by P is defined as $\|u\|_P = \sqrt{\langle u, P u \rangle}$ for all $u \in \mathbb{R}^d$. We denote with $\mathcal{D}_\eta \subseteq \mathcal{S}_+(\mathbb{R}^d)$ the set of matrices whose eigenvalues belong to the interval $[\eta, \infty[$ and $\mathcal{D}_\eta^\mu \subseteq \mathcal{S}_+(\mathbb{R}^d)$ the set of matrices whose eigenvalues belong to the interval $[\eta, \mu]$. If $P \in \mathcal{D}_\eta^\mu$, then we have

$$\eta \|u\|^2 \leq \|u\|_P^2 \leq \mu \|u\|^2, \quad \forall u \in \mathbb{R}^d, \tag{6}$$

and likewise

$$\mu^{-1} \|u\|^2 \leq \|u\|_{P^{-1}}^2 \leq \eta^{-1} \|u\|^2, \quad \forall u \in \mathbb{R}^d. \tag{7}$$

Finally, we recall that the relative interior of $\Omega \subseteq \mathbb{R}^d$ is the set $\text{relint}(\Omega) = \{u \in \Omega : \exists \epsilon > 0 \text{ s.t. } B(u, \epsilon) \cap \text{aff}(\Omega) \subseteq \Omega\}$, where $B(u, \epsilon)$ is the ball of center u and radius ϵ , and $\text{aff}(\Omega)$ is the affine hull of Ω .

2.1 Basic definitions and results

Definition 1 A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to have a L -Lipschitz continuous gradient if the following property holds

$$\|\nabla f(\tilde{u}) - \nabla f(\bar{u})\| \leq L\|\tilde{u} - \bar{u}\|, \quad \forall \tilde{u}, \bar{u} \in \mathbb{R}^d.$$

For our purposes, it is useful to recall the following variant of the well-known Descent Lemma, where the standard Euclidean norm is replaced by the norm induced by a symmetric positive definite matrix.

Lemma 1 [8, Lemma 6] Let $f : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ be a continuously differentiable function with L -Lipschitz continuous gradient. If $P \in \mathcal{D}_\eta$ and $\bar{L} \geq L/\eta$, we have

$$f(\tilde{u}) \leq f(\bar{u}) + \langle \nabla f(\bar{u}), \tilde{u} - \bar{u} \rangle + \frac{\bar{L}}{2}\|\tilde{u} - \bar{u}\|_P^2, \quad \forall \tilde{u}, \bar{u} \in \mathbb{R}^d. \tag{8}$$

Definition 2 [39, p. 104] Given a proper, convex, lower semicontinuous function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$, the convex conjugate of φ is the function

$$\varphi^* : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}, \quad \varphi^*(w) = \sup_{u \in \mathbb{R}^d} \langle w, u \rangle - \varphi(u), \quad \forall w \in \mathbb{R}^d.$$

The following well-known result holds for the biconjugate function $(\varphi^*)^*$.

Lemma 2 [39, Theorem 12.2] Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ be proper, convex, and lower semicontinuous. Then φ^* is convex and lower semicontinuous and $(\varphi^*)^* = \varphi$, namely,

$$\varphi(u) = \sup_{w \in \mathbb{R}^d} \langle u, w \rangle - \varphi^*(w), \quad \forall u \in \mathbb{R}^d.$$

Definition 3 [40, p. 278], [10, p. 877] The proximal operator of a proper, convex, lower semicontinuous function φ with parameter $\gamma > 0$ is the map $\text{prox}_{\gamma\varphi} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined as

$$\text{prox}_{\gamma\varphi}(a) = \arg \min_{u \in \mathbb{R}^d} \varphi(u) + \frac{1}{2\gamma}\|u - a\|^2, \quad \forall a \in \mathbb{R}^d.$$

Similarly, we define the proximal operator with respect to the norm induced by $P \in \mathcal{D}_\eta$ as

$$\text{prox}_{\gamma\varphi}^P(a) = \arg \min_{u \in \mathbb{R}^d} \varphi(u) + \frac{1}{2\gamma}\|u - a\|_P^2, \quad \forall a \in \mathbb{R}^d.$$

The following result holds for proximal points evaluated at perturbed points.

Lemma 3 [32, Lemma 3.3] *Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ be proper, convex, and lower semicontinuous, and $x, e \in \mathbb{R}^d$. Then the equality $y = \text{prox}_\varphi(x + e)$ is equivalent to the following inequality:*

$$\|y - z\|^2 \leq \|x - z\|^2 - \|x - y\|^2 + 2\langle y - z, e \rangle + 2\varphi(z) - 2\varphi(y), \quad \forall z \in \mathbb{R}^d.$$

Finally, the following three technical lemmas will be required in order to derive the convergence analysis of our proposed algorithm.

Lemma 4 [19, Lemma 1] *Let $\{a_n\}_{n \in \mathbb{N}}, \{b_n\}_{n \in \mathbb{N}}, \{c_n\}_{n \in \mathbb{N}}$ be sequences of real nonnegative numbers, with $\{b_n\}_{n \in \mathbb{N}}$ being a monotone nondecreasing sequence, satisfying the following recursive property*

$$a_n^2 \leq b_n + \sum_{k=1}^n c_k a_k, \quad \forall n \geq 1. \tag{9}$$

Then the following inequality holds:

$$a_n \leq \frac{1}{2} \sum_{k=1}^n c_k + \left(b_n + \left(\frac{1}{2} \sum_{k=1}^n c_k \right)^2 \right)^{\frac{1}{2}}, \quad \forall n \geq 1. \tag{10}$$

Lemma 5 [41] *Let $\{a_n\}_{n \in \mathbb{N}}, \{b_n\}_{n \in \mathbb{N}}$ and $\{c_n\}_{n \in \mathbb{N}}$ be sequences of real nonnegative numbers such that $a_{n+1} \leq (1 + b_n)a_n + c_n$ and $\sum_{n=0}^\infty b_n < \infty, \sum_{n=0}^\infty c_n < \infty$. Then, the sequence $\{a_n\}_{n \in \mathbb{N}}$ converges.*

Lemma 6 [42, Lemma 2.3] *Let $\{P_n\}_{n \in \mathbb{N}} \subseteq \mathcal{D}_\eta^\mu$ be a sequence of scaling matrices satisfying Assumption 2(ii). Then, there exists $P \in \mathcal{D}_\eta$ such that $\lim_{n \rightarrow \infty} P_n = P$ pointwise.*

2.2 Problem formulation

From now on, we consider the optimization problem (1) under the following assumptions.

- Assumption 1** (i) $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is continuously differentiable with L –Lipschitz continuous gradient.
- (ii) $h : \mathbb{R}^{d'} \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous and convex.
- (iii) $W \in \mathbb{R}^{d' \times d}$ and there exists $u_0 \in \mathbb{R}^d$ such that $Wu_0 \in \text{reint}(\text{dom}(h))$.
- (iv) Problem (1) admits at least one solution $\hat{u} \in \mathbb{R}^d$.

We remark that the assumption on W is needed to guarantee that the subdifferential rule $\partial(h \circ W)(u) = W^T \partial h(Wu)$ holds, so that we can interpret the minimum points of (1) as solutions of appropriate variational equations, as stated below.

Lemma 7 [32, Lemma 3.1] *Under Assumption 1, a point $\hat{u} \in \mathbb{R}^d$ is a solution of problem (1) if and only if the following conditions hold*

$$\begin{cases} \nabla f(\hat{u}) + W^T \hat{v} = 0 \\ \hat{v} = \text{prox}_{\beta\alpha^{-1}h^*}(\hat{v} + \beta\alpha^{-1}W\hat{u}), \end{cases} \quad \forall \alpha, \beta > 0. \tag{11}$$

Due to Lemma 2, we can equivalently reformulate the convex problem (1) as the following convex-concave saddle-point problem

$$\min_{u \in \mathbb{R}^d} \max_{v \in \mathbb{R}^{d'}} \mathcal{L}(u, v) \equiv f(u) + \langle Wu, v \rangle - h^*(v), \tag{12}$$

where $\mathcal{L}(u, v)$ denotes the primal–dual function. A solution of (12) is any point $(\hat{u}, \hat{v}) \in \mathbb{R}^d \times \mathbb{R}^{d'}$ such that

$$\mathcal{L}(\hat{u}, v) \leq \mathcal{L}(\hat{u}, \hat{v}) \leq \mathcal{L}(u, \hat{v}), \quad \forall u \in \mathbb{R}^d, \forall v \in \mathbb{R}^{d'}. \tag{13}$$

2.3 Inexact proximal evaluations via primal–dual iterates

In this section, we show that the proximal operator with a metric defined by a symmetric positive definite matrix P can be approximated by an appropriate sequence of primal–dual iterates. The following result generalizes the one derived in [34, Theorem 1] by taking into account the presence of the scaling matrix.

Lemma 8 *Suppose that $h : \mathbb{R}^{d'} \rightarrow \mathbb{R} \cup \{\infty\}$ and $W \in \mathbb{R}^{d' \times d}$ satisfy Assumption 1(ii–iii). Let $P \in \mathbb{R}^{d \times d}$ be a symmetric positive definite matrix and $a \in \mathbb{R}^d$. Choose $\alpha > 0, 0 < \beta < 2/\|WP^{-1}W^T\|, v^0 \in \mathbb{R}^{d'}$, define the sequence*

$$v^{k+1} = \text{prox}_{\beta\alpha^{-1}h^*}(v^k + \beta\alpha^{-1}W(a - \alpha P^{-1}W^T v^k)), \quad \forall k \geq 0, \tag{14}$$

and its limit $\hat{v} = \lim_{k \rightarrow \infty} v^k$. Then we have

$$\text{prox}_{\alpha h \circ W}^P(a) = a - \alpha P^{-1}W^T \hat{v}. \tag{15}$$

Proof Let $\hat{a} = \text{prox}_{\alpha h \circ W}^P(a)$, which is equivalent to writing

$$\hat{a} = \underset{u \in \mathbb{R}^d}{\text{argmin}} \frac{1}{2\alpha} \|u - a\|_P^2 + h(Wu).$$

By applying Lemma 7 to the above minimization problem, we obtain the following system of variational equations

$$\hat{a} = \text{prox}_{\alpha h \circ W}^P(a) \iff \begin{cases} P(\hat{a} - a) + \alpha W^T \hat{v} = 0 \\ \hat{v} = \text{prox}_{\beta\alpha^{-1}h^*}(\hat{v} + \beta\alpha^{-1}W\hat{a}) \end{cases} \iff \begin{cases} \hat{a} = a - \alpha P^{-1}W^T \hat{v} \\ \hat{v} = \text{prox}_{\beta\alpha^{-1}h^*}(\hat{v} + \beta\alpha^{-1}W\hat{a}). \end{cases} \tag{16}$$

Consider the sequences $\{a^k\}_{k \in \mathbb{N}}$ and $\{v^k\}_{k \in \mathbb{N}}$ (with arbitrary v^0) given by

$$\begin{cases} a^k = a - \alpha P^{-1}W^T v^k \\ v^{k+1} = \text{prox}_{\beta\alpha^{-1}h^*}(v^k + \beta\alpha^{-1}W a^k) \end{cases}, \quad k = 0, 1, \dots \tag{17}$$

or equivalently

$$v^{k+1} = \text{prox}_{\beta\alpha^{-1}h^*}(v^k + \beta\alpha^{-1}W(a - \alpha P^{-1}W^T v^k)), \quad k = 0, 1, \dots$$

In virtue of (16), the sequence $\{v^k\}_{k \in \mathbb{N}}$ can be interpreted as a fixed-point iteration applied to the operator

$$T(v) = \text{prox}_{\beta\alpha^{-1}h^*}(v + \beta\alpha^{-1}W(a - \alpha P^{-1}W^T v)). \tag{18}$$

By Banach fixed-point theorem, the sequence $\{v^k\}_{k \in \mathbb{N}}$ converges to a fixed-point of T provided that T is a contraction. Since the operator $\text{prox}_{\beta\alpha^{-1}h^*}$ is non-expansive, we have

$$\begin{aligned} \|T(v) - T(u)\| &\leq \|v - u + \beta\alpha^{-1}(\alpha W P^{-1}W^T u - \alpha W P^{-1}W^T v)\| \\ &= \|(I - \beta W P^{-1}W^T)(v - u)\| \\ &\leq \|I - \beta W P^{-1}W^T\| \|v - u\|, \end{aligned}$$

so T is a contraction as long as $\|I - \beta W P^{-1}W^T\| < 1$. Since P is symmetric, it holds $\|I - \beta W P^{-1}W^T\| = \rho(I - \beta W P^{-1}W^T)$, where $\rho(\cdot)$ denotes the spectral radius of a matrix, and hence T is a contraction if and only if $0 < \beta < 2/\|W P^{-1}W^T\|$. Therefore, $\{v^k\}_{k \in \mathbb{N}}$ converges to a fixed-point \hat{v} of T , and by continuity, the sequence a^k also converges to the point $\hat{a} = a - \alpha P^{-1}W^T \hat{v}$. Thus (16) yields the thesis. \square

We remark that, although Lemma 8 will not be employed in the convergence analysis, it will enable us to interpret our proposed method as a particular variable metric proximal–gradient method with extrapolation of the form (3), where the proximal operator is computed through the primal–dual procedure (14). We refer the reader to the upcoming Sect. 3.1 for more details on this issue.

Algorithm 1 Nested Primal–Dual Variable Metric method

Choose $k_{\max} \in \mathbb{N}$, $u_{-1} \in \mathbb{R}^d$, $u_0 = u_{-1}$, $v_{-1}^{k_{\max}} \in \mathbb{R}^{d'}$, $\epsilon \in (0, 1)$, $\delta \in (0, 1)$, $L_{-1} > 0$, $\eta > 0$, $\alpha_{-1} = \epsilon/L_{-1}$, $P_{-1} \in \mathcal{D}_\eta$, $\beta_{-1} = \epsilon/(\|P_{-1}^{-1}\| \|W\|^2)$.

FOR $n = 0, 1, \dots$

1. Choose $\gamma_n \geq 0$ and compute the extrapolated point

$$\bar{u}_n = u_n + \gamma_n(u_n - u_{n-1}). \tag{19}$$

2. Choose $P_n \in \mathcal{D}_\eta$.

3. Set $i_n = 0$, $L_n = L_{n-1}$, $\alpha_n = \epsilon/L_n$, $\beta_n = \epsilon/(\|P_n^{-1}\| \|W\|^2)$, $v_n^0 = v_{n-1}^{k_{\max}}$.

4. Compute k_{\max} primal–dual iterates:

FOR $k = 0, 1, \dots, k_{\max} - 1$

$$u_n^k = \bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n) - \alpha_n P_n^{-1} W^T v_n^k \tag{20}$$

$$v_n^{k+1} = \text{prox}_{\beta_n \alpha_n^{-1} h^*}(v_n^k + \beta_n \alpha_n^{-1} W u_n^k). \tag{21}$$

5. Compute $u_n^{k_{\max}} = \bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n) - \alpha_n P_n^{-1} W^T v_n^{k_{\max}}$.

6. Compute

$$\tilde{u}_n = \frac{1}{k_{\max}} \sum_{k=1}^{k_{\max}} u_n^k. \tag{22}$$

7. If $f(\tilde{u}_n) \leq f(\bar{u}_n) + \nabla f(\bar{u}_n)^T (\tilde{u}_n - \bar{u}_n) + \frac{L_n}{2} \|\tilde{u}_n - \bar{u}_n\|_{P_n}^2$

Go to Step 8.

ELSE

Set $i_n = i_n + 1$, $L_n = L_{n-1}/\delta^{i_n}$, $\alpha_n = \epsilon/L_n$ and go to Step 4.

8. Set $u_{n+1} = \tilde{u}_n$.
-

3 A nested primal–dual variable metric method

In this section, we propose and analyse our nested primal–dual variable metric method for solving problem (1). The resulting scheme can be considered as a variable metric proximal–gradient algorithm with extrapolation, where the proximal operator is approximated by means of a prefixed number of primal–dual steps. Note that our proposed method generalizes the one in [34], by including a backtracking procedure for the steplength and the selection of a variable scaling matrix.

3.1 The proposed method

We report our proposed method in Algorithm 1. It requires the choice of the prefixed number of primal–dual iterates $k_{\max} \in \mathbb{N}$, the initial guesses $u_0 = u_{-1} \in \mathbb{R}^d$, $v_{-1}^{k_{\max}} \in \mathbb{R}^{d'}$.

$\mathbb{R}^{d'}$, the parameters $\epsilon, \delta \in (0, 1)$, an approximation of the Lipschitz constant $L_{-1} > 0$ and the metric parameters $\eta > 0, \alpha_{-1} = \epsilon/L_{-1}, P_{-1} \in \mathcal{D}_\eta$.

In Step 1, we choose the extrapolation parameter $\gamma_n \geq 0$ and compute the extrapolated iterate \bar{u}_n according to (19), whereas Step 2 is devoted to the choice of the scaling matrix $P_n \in \mathcal{D}_\eta$.

At Step 3, we initialize the approximation of the Lipschitz constant L_n as the value computed at the previous iteration, the primal steplength as $\alpha_n = \epsilon/L_n$, the dual steplength as $\beta_n = \epsilon/\|P_n^{-1}\| \|W\|^2$, and the initial dual iterate as $v_n^0 = v_{n-1}^{k_{\max}}$, i.e., the inner primal–dual loop is *warm-started* with the outcome of the loop at the previous iteration. Such a warm-start strategy is borrowed from the works [32, 34] and is crucial to guarantee the convergence of the algorithm.

Steps 4–7 define the backtracking procedure that is needed for adaptively computing the approximation of the Lipschitz constant L_n , the corresponding steplength α_n and the next iterate u_{n+1} . First, we compute k_{\max} primal–dual iterates according to the procedure discussed in Sect. 2.3 (Step 4). More precisely, equations (20)–(21) represent k_{\max} iterations of the primal–dual method (14) with $a = \bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n)$, $\alpha = \alpha_n$, $P = P_n$ and $\beta = \beta_n$; then, in virtue of Lemma 8, the sequence $\{u_n^k\}_{k=0}^{k_{\max}-1}$ can be considered as approximating the proximal–gradient point $\text{prox}_{\alpha_n h \circ W}^{P_n}(\bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n))$. Next, we compute an additional primal iterate $u_n^{k_{\max}}$ (Step 5) and average the primal iterates $\{u_n^k\}_{k=1}^{k_{\max}}$ over the number of inner iterations (Step 6). Finally, we check a backtracking condition based on Lemma 1 (Step 7); if the condition is satisfied, then we accept the average of the primal iterates as the next iterate (Step 8), otherwise we increase L_n by a factor $1/\delta$ and repeat the backtracking procedure.

Remark 1 Note that the backtracking procedure at Steps 4–7 terminates in a finite number of steps, as the backtracking condition at Step 7 is accepted whenever $L_n \geq L/\eta$ (see Lemma 1). Furthermore, the sequence $\{\alpha_n\}_{n \in \mathbb{N}}$ is nonincreasing thanks to Step 3 and the backtracking procedure at Steps 4–7. Note that $\{\alpha_n\}_{n \in \mathbb{N}}$ is also bounded away from zero: indeed, either $L_n < L/\eta$ and hence $\alpha_n > \epsilon\eta/L$ for all $n \geq 0$, or otherwise let $n^* \geq 0$ be the first iteration such that $L_{n^*} \geq L/\eta$; in the latter case, if $n^* = 0$, then Lemma 1 and Steps 7–8 of Algorithm 1 imply $\alpha_n = \alpha_0$ for all $n \geq 0$, otherwise it must be $\delta L_{n^*} < L/\eta$ and thus $\alpha_n \geq \alpha_{n^*} > (\epsilon\delta\eta)/L$ for all $n \geq 0$. In conclusion, the following inequalities hold:

$$0 < \min \left\{ \alpha_0, \frac{\epsilon\delta\eta}{L} \right\} \leq \alpha_n \leq \alpha_{n-1} \leq \alpha_{-1}, \quad \forall n \geq 0. \tag{23}$$

3.2 Convergence analysis

In this section, we carry out the convergence analysis for Algorithm 1. Under appropriate conditions on the extrapolation parameter and the scaling matrix, we show that the iterates sequence $\{u_n\}_{n \in \mathbb{N}}$ converge to a solution of problem (1).

For our purposes, we define the sequence of matrices $\{D_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}^{d' \times d'}$ as

$$D_n = I_{d'} - \beta_n W P_n^{-1} W^T, \quad \forall n \geq 0. \tag{24}$$

Note that, by construction, the matrix D_n are real and symmetric since $P_n \in \mathcal{D}_\eta$. Moreover, we can prove the following.

Lemma 9 *Let $\{D_n\}_{n \in \mathbb{N}}$ be the sequence of matrices defined as in (24). Then, for all $n \geq 0$, D_n is a real symmetric positive definite matrix and its eigenvalues are bounded away from zero by a constant independent of n .*

Proof We just need to prove that D_n is positive definite and, since it is symmetric, it is sufficient to show that $\lambda_{\min}(D_n) > 0$. In this respect, we have that

$$\begin{aligned} \lambda_{\min}(D_n) &= 1 - \beta_n \lambda_{\max}(W P_n^{-1} W^T) \\ &= 1 - \beta_n \|W P_n^{-1} W^T\| \\ &\geq 1 - \beta_n \|P_n^{-1}\| \|W\|^2 \\ &= 1 - \epsilon > 0, \quad \forall n \geq 0, \end{aligned} \quad (25)$$

where the first inequality follows from the submultiplicative property of the spectral norm and (7), and the last equality is due to the choice of β_n in Algorithm 1. The final result holds because $\epsilon \in (0, 1)$. \square

Based on the previous result, we can consider the norm induced by D_n , i.e.,

$$\begin{aligned} \|v\|_{D_n}^2 &= \|(I - \beta_n W P_n^{-1} W^T)^{\frac{1}{2}} v\|^2 \\ &= v^T (I - \beta_n W P_n^{-1} W^T) v \\ &= \|v\|^2 - \beta_n \|P_n^{-\frac{1}{2}} W^T v\|^2, \quad \forall v \in \mathbb{R}^{d'}. \end{aligned} \quad (26)$$

The following lemma contains some crucial descent inequalities involving the primal–dual function \mathcal{L} and the iterates generated by Algorithm 1. The proof of this result is similar to the one of [34, Lemma 7], although some modifications are needed in order to address the presence of the variable metric and the backtracking procedure on the steplength α_n , which were absent in [34]. Unlike in [34], the inequalities are not given for a generic primal–dual sequence, rather they specifically hold at the iterates obtained by averaging the primal and dual inner iterates of Algorithm 1.

Lemma 10 *Suppose that Assumption 1 holds. Let $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ be the primal–dual sequence generated by Algorithm 1, and let $(\hat{u}, \hat{v}) \in \mathbb{R}^d \times \mathbb{R}^{d'}$ be a solution of the primal–dual problem (12).*

(i) *Define the sequence of dual iterates $\{\tilde{v}_n\}_{n \in \mathbb{N}}$ as*

$$\tilde{v}_n = \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} v_n^{k+1}, \quad \forall n \geq 0. \quad (27)$$

Then, for all $n \geq 0$ and for all $u' \in \mathbb{R}^d$, we have

$$\begin{aligned} &\mathcal{L}(u_{n+1}, \tilde{v}_n) + \frac{1}{2\alpha_n} \|u_{n+1} - u'\|_{P_n}^2 \\ &\leq \mathcal{L}(u', \tilde{v}_n) + \frac{1}{2\alpha_n} \|\bar{u}_n - u'\|_{P_n}^2 - \frac{1}{2} \left(\frac{1}{\alpha_n} - L_n \right) \|u_{n+1} - \bar{u}_n\|_{P_n}^2. \end{aligned} \tag{28}$$

(ii) For all $n \geq 0$ and for all $v' \in \mathbb{R}^{d'}$, we have

$$\begin{aligned} &\mathcal{L}(u_{n+1}, v') + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_{n+1}^0 - v'\|_{D_n}^2 \\ &\leq \mathcal{L}(u_{n+1}, \tilde{v}_n) + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_n^0 - v'\|_{D_n}^2 - \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2. \end{aligned} \tag{29}$$

(iii) For all $n \geq 0$, for all $u' \in \mathbb{R}^d$ and $v' \in \mathbb{R}^{d'}$, we have

$$\begin{aligned} &\mathcal{L}(u_{n+1}, v') + \frac{1}{2\alpha_n} \|u_{n+1} - u'\|_{P_n}^2 + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_{n+1}^0 - v'\|_{D_n}^2 \\ &\leq \mathcal{L}(u', \tilde{v}_n) + \frac{1}{2\alpha_n} \|\bar{u}_n - u'\|_{P_n}^2 + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_n^0 - v'\|_{D_n}^2 \\ &\quad - \frac{1}{2} \left(\frac{1}{\alpha_n} - L_n \right) \|u_{n+1} - \bar{u}_n\|_{P_n}^2 - \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2. \end{aligned} \tag{30}$$

Proof See Appendix A. □

We now discuss the assumptions needed on the inertial parameters $\{\gamma_n\}_{n \in \mathbb{N}}$ and the scaling matrices $\{P_n\}_{n \in \mathbb{N}}$ in order to ensure the convergence of Algorithm 1.

Assumption 2 (i) The sequence of inertial parameters $\{\gamma_n\}_{n \in \mathbb{N}}$ of Algorithm 1 complies with the following condition

$$\sum_{n=0}^{\infty} \gamma_n \|u_n - u_{n-1}\| < \infty. \tag{31}$$

(ii) The sequence of scaling matrices $\{P_n\}_{n \in \mathbb{N}}$ of Algorithm 1 is chosen so that

$$P_n \leq (1 + \zeta_{n-1})P_{n-1}, \quad \forall n \geq 0, \quad \text{where } \zeta_{n-1} \geq 0, \quad \sum_{n=0}^{\infty} \zeta_{n-1} < \infty. \tag{32}$$

Remark 2 Condition (31) has been previously employed in [34] for proving the convergence of a special instance of Algorithm 1 devoid of variable metrics and backtracking procedures. It can be easily implemented in practice, as it depends only on the past

iterates u_n, u_{n-1} . As observed in [34], (31) has the practical effect of “shrinking” the inertial parameter when the iteration number n is large.

In a similar fashion, condition (32) is forcing the matrices $\{P_n\}_{n \in \mathbb{N}}$ to converge to a symmetric positive definite matrix at a sufficiently fast rate controlled by the parameters sequence $\{\zeta_n\}_{n \in \mathbb{N}}$. Such a condition can be easily enforced if one selects the scaling matrix as a diagonal matrix, computed according to a specific update rule such as the Majorization–Minimization strategy or the Split-Gradient decomposition [8, 9, 28], and then constrains its elements to an interval of diminishing size. In section 4.2 we show that the preconditioning matrices (5) naturally comply with this assumption.

Remark 3 If condition (32) is assumed to be satisfied, then [42, Lemma 2.1] and (7) yield the following matrix relations

$$((1 + \zeta_{n-1})P_{n-1})^{-1} \leq P_n^{-1} \leq \frac{1}{\eta} I_d. \tag{33}$$

Since the spectral norm is monotone [43, Ex. 2.2-10], the above relation implies

$$\eta \leq \frac{1}{\|P_n^{-1}\|} \leq \frac{1 + \zeta_{n-1}}{\|P_{n-1}^{-1}\|}.$$

Consequently, the parameters $\{\beta_n\}_{n \in \mathbb{N}}$ satisfy the following inequalities

$$0 < \frac{\epsilon \eta}{\|W\|^2} \leq \beta_n \leq (1 + \zeta_{n-1})\beta_{n-1}, \quad \forall n \geq 0. \tag{34}$$

Remark 4 Regarding the sequence of matrices $\{D_n\}_{n \in \mathbb{N}}$, we derive the following inequalities

$$\begin{aligned} \|v\|_{D_n}^2 &= \|v\|^2 - \beta_n \|W^T v\|_{P_n^{-1}}^2 \\ &\leq \|v\|^2 - \frac{\beta_n}{1 + \zeta_{n-1}} \|W^T v\|_{P_{n-1}^{-1}}^2 \\ &= \frac{(1 + \zeta_{n-1})\|v\|^2 - \beta_n \|W^T v\|_{P_{n-1}^{-1}}^2}{1 + \zeta_{n-1}} \\ &= \frac{\|v\|_{D_{n-1}}^2 + \zeta_{n-1}\|v\|^2 + (\beta_{n-1} - \beta_n)\|W^T v\|_{P_{n-1}^{-1}}^2}{1 + \zeta_{n-1}} \\ &\leq \|v\|_{D_{n-1}}^2 + \zeta_{n-1}\|v\|^2 + \frac{(\beta_{n-1} - \beta_n)\|W\|^2}{\eta} \|v\|^2 \\ &\leq \left(1 + \frac{\zeta_{n-1}}{1 - \epsilon} + \frac{(\beta_{n-1} - \beta_n)\|W\|^2}{\eta(1 - \epsilon)}\right) \|v\|_{D_{n-1}}^2, \end{aligned}$$

where the first inequality follows from (33) and the last one is due to the combination of (6) and (25). In conclusion, there exists a sequence $\{\tilde{\zeta}_{n-1}\}_{n \in \mathbb{N}}$ such that

$$D_n \leq \left(1 + \tilde{\zeta}_{n-1}\right) D_{n-1}, \quad \forall n \geq 0, \quad \text{where } \tilde{\zeta}_{n-1} \geq \zeta_{n-1} \geq 0, \quad \sum_{n=0}^{\infty} \tilde{\zeta}_{n-1} < \infty. \tag{35}$$

The main convergence result for Algorithm 1 is stated below. The line of proof employed is analogous to the one in [34, Theorem 2], even though it must be adapted to the presence of variable metrics and variable steplengths.

Theorem 1 *Suppose that Assumption 1 holds and let $(\hat{u}, \hat{v}) \in \mathbb{R}^d \times \mathbb{R}^{d'}$ be a solution of the primal-dual problem (12). Let $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ be the primal-dual sequence generated by Algorithm 1. Suppose that the inertial parameters $\{\gamma_n\}_{n \in \mathbb{N}}$ and scaling matrices $\{P_n\}_{n \in \mathbb{N}}$ of Algorithm 1 comply with Assumption 2. Then the following statements hold true.*

- (i) *The sequence $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ is bounded.*
- (ii) *Given a saddle point (\hat{u}, \hat{v}) solution of (12), the sequence $\{\beta_{n-1} k_{\max} \|\hat{u} - u_n\|_{P_{n-1}}^2 + \alpha_{n-1}^2 \|\hat{v} - v_n^0\|_{D_{n-1}}^2\}_{n \in \mathbb{N}}$ converges.*
- (iii) *The sequence $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ converges to a solution of (12).*

Before going into the technical details of the proof, let us briefly summarize how we intend to prove each item of Theorem 1.

- (i) The proof of item (i) relies on an upper bound on the quantity $\|u_{n+1} - \hat{u}\|_{P_n}$, which is obtained by employing Lemma 4 in combination with Assumption 2.
- (ii) Item (ii) follows by applying Lemma 5 to the sequence $\{\beta_{n-1} k_{\max} \|\hat{u} - u_n\|_{P_{n-1}}^2 + \alpha_{n-1}^2 \|\hat{v} - v_n^0\|_{D_{n-1}}^2\}_{n \in \mathbb{N}}$, which is possible thanks to item (i).
- (iii) For item (iii), we first show that any limit point $(u^\dagger, v^\dagger) \in \mathbb{R}^d \times \mathbb{R}^{d'}$ of the sequence $\{(u^n, v_n^0)\}_{n \in \mathbb{N}}$ is a solution to problem (12); then, the thesis follows by applying item (ii) with the choice $(\hat{u}, \hat{v}) = (u^\dagger, v^\dagger)$.

Proof (i) Consider inequality (30) with $u' = \hat{u}$, $v' = \hat{v}$ and the term $-\frac{1}{2} \left(\frac{1}{\alpha_n} - L_n\right) \|u_{n+1} - \bar{u}_n\|_{P_n}^2 - \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2$ discarded, which amounts to writing

$$\begin{aligned} \mathcal{L}(u_{n+1}, \hat{v}) &+ \frac{1}{2\alpha_n} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \\ &\leq \mathcal{L}(\hat{u}, \tilde{v}_n) + \frac{1}{2\alpha_n} \|\bar{u}_n - \hat{u}\|_{P_n}^2 + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_n^0 - \hat{v}\|_{D_n}^2. \end{aligned}$$

By observing that $\mathcal{L}(u_{n+1}, \hat{v}) \geq \mathcal{L}(\hat{u}, \tilde{v}_n)$, due to the fact that (\hat{u}, \hat{v}) is a solution of the primal-dual problem (12), and multiplying the above inequality by $\alpha_n \beta_n$, we deduce the inequality

$$\frac{\beta_n}{2} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \frac{\alpha_n^2}{2k_{\max}} \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \leq \frac{\beta_n}{2} \|\bar{u}_n - \hat{u}\|_{P_n}^2 + \frac{\alpha_n^2}{2k_{\max}} \|v_n^0 - \hat{v}\|_{D_n}^2. \tag{36}$$

By recalling that the sequences $\{\alpha_n\}_{n \in \mathbb{N}}$ and $\{\beta_n\}_{n \in \mathbb{N}}$ satisfy (23) and (34), respectively, and applying conditions (32) and (35) to inequality (36), we get

$$\begin{aligned} & \frac{\beta_n}{2} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \frac{\alpha_n^2}{2k_{\max}} \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \\ & \leq (1 + \tilde{\zeta}_{n-1})(1 + \zeta_{n-1}) \left(\frac{\beta_{n-1}}{2} \|\bar{u}_n - \hat{u}\|_{P_{n-1}}^2 + \frac{\alpha_{n-1}^2}{2k_{\max}} \|v_n^0 - \hat{v}\|_{D_{n-1}}^2 \right) \\ & = (1 + \chi_{n-1}) \left(\frac{\beta_{n-1}}{2} \|\bar{u}_n - \hat{u}\|_{P_{n-1}}^2 + \frac{\alpha_{n-1}^2}{2k_{\max}} \|v_n^0 - \hat{v}\|_{D_{n-1}}^2 \right), \end{aligned} \tag{37}$$

where $\chi_{n-1} = \zeta_{n-1} + \tilde{\zeta}_{n-1} + \zeta_{n-1}\tilde{\zeta}_{n-1}$ for all $n \geq 0$ and $\{\chi_n\}_{n \in \mathbb{N}}$ is still a nonnegative summable sequence. Since $\bar{u}_n = u_n + \gamma_n(u_n - u_{n-1})$, an application of the Cauchy-Schwarz yields

$$\begin{aligned} & \frac{\beta_n}{2} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \frac{\alpha_n^2}{2k_{\max}} \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \\ & \leq (1 + \chi_{n-1}) \left(\frac{\beta_{n-1}}{2} \|u_n - \hat{u}\|_{P_{n-1}}^2 + \frac{\alpha_{n-1}^2}{2k_{\max}} \|v_n^0 - \hat{v}\|_{D_{n-1}}^2 \right) \\ & + \beta_{n-1}(1 + \chi_{n-1}) \left(\frac{\gamma_n^2}{2} \|u_n - u_{n-1}\|_{P_{n-1}}^2 + \gamma_n \langle u_n - \hat{u}, P_{n-1}(u_n - u_{n-1}) \rangle \right) \\ & \leq (1 + \chi_{n-1}) \left(\frac{\beta_{n-1}}{2} \|u_n - \hat{u}\|_{P_{n-1}}^2 + \frac{\alpha_{n-1}^2}{2k_{\max}} \|v_n^0 - \hat{v}\|_{D_{n-1}}^2 \right) \\ & + \beta_{n-1}(1 + \chi_{n-1}) \left(\frac{\gamma_n^2}{2} \|u_n - u_{n-1}\|_{P_{n-1}}^2 + \gamma_n \|u_n - \hat{u}\|_{P_{n-1}} \|u_n - u_{n-1}\|_{P_{n-1}} \right). \end{aligned} \tag{38}$$

By applying recursively inequality (38), we get

$$\begin{aligned} & \frac{\beta_n}{2} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \frac{\alpha_n^2}{2k_{\max}} \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \\ & \leq \left(\prod_{k=0}^n (1 + \chi_{k-1}) \right) \left(\frac{\beta_{-1}}{2} \|u_0 - \hat{u}\|_{P_{-1}}^2 + \frac{\alpha_{-1}^2}{2k_{\max}} \|v_0^0 - \hat{v}\|_{D_{-1}}^2 \right) \\ & + \sum_{k=0}^n \beta_{k-1} \left(\prod_{i=k}^n (1 + \chi_{i-1}) \right) \left(\frac{\gamma_k^2}{2} \|u_k - u_{k-1}\|_{P_{k-1}}^2 + \gamma_k \|u_k - \hat{u}\|_{P_{k-1}} \|u_k - u_{k-1}\|_{P_{k-1}} \right). \end{aligned}$$

Now, if we consider the sequence

$$\Lambda_n = \prod_{k=0}^n (1 + \chi_{k-1}), \quad n \geq 0,$$

we note that $\Lambda_n = (1 + \chi_{n-1})\Lambda_{n-1}$, which implies that $\{\Lambda_n\}_{n \in \mathbb{N}}$ is an increasing and convergent sequence, see Lemma 5. By setting $\Lambda = \lim_{n \rightarrow \infty} \Lambda_n$, it follows that $\Lambda_n \leq \Lambda$ for all $n \geq 0$, $\prod_{i=k}^n (1 + \chi_{i-1}) \leq \Lambda$ for $0 \leq k \leq n$, and the previous inequality yields

$$\begin{aligned} & \frac{\beta_n}{2} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \frac{\alpha_n^2}{2k_{\max}} \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \leq \Lambda \left(\frac{\beta_{-1}}{2} \|u_0 - \hat{u}\|_{P_{-1}}^2 + \frac{\alpha_{-1}^2}{2k_{\max}} \|v_0^0 - \hat{v}\|_{D_{-1}}^2 \right) \\ & + \sum_{k=0}^n \beta_{k-1} \Lambda \left(\frac{\gamma_k^2}{2} \|u_k - u_{k-1}\|_{P_{k-1}}^2 + \gamma_k \|u_k - \hat{u}\|_{P_{k-1}} \|u_k - u_{k-1}\|_{P_{k-1}} \right). \end{aligned} \tag{39}$$

By discarding the term proportional to $\|v_{n+1}^0 - \hat{v}\|_{D_n}^2$ on the left-hand side of the previous inequality, multiplying both sides by a factor $2/\beta_n$, using the inequalities $(\epsilon\eta)/\|W\|^2 \leq \beta_k \leq (1 + \zeta_{k-1})\beta_{k-1} \leq \Lambda_k \beta_{-1} \leq \Lambda \beta_{-1}$ for $k = 0, \dots, n$, and adding the term $\frac{2\Lambda^2\|W\|^2}{\epsilon\eta} \gamma_{n+1} \|u_{n+1} - \hat{u}\|_{P_n} \|u_{n+1} - u_n\|_{P_n}$ to the right-hand side, one gets

$$\begin{aligned} \|u_{n+1} - \hat{u}\|_{P_n}^2 & \leq \frac{\Lambda\|W\|^2}{\epsilon\eta} \left(\beta_{-1} \|u_0 - \hat{u}\|_{P_{-1}}^2 + \frac{\alpha_{-1}^2}{k_{\max}} \|v_0^0 - \hat{v}\|_{D_{-1}}^2 + \sum_{k=0}^n \Lambda \beta_{-1} \gamma_k^2 \|u_k - u_{k-1}\|_{P_{k-1}}^2 \right) \\ & + \sum_{k=0}^{n+1} \frac{2\Lambda^2\|W\|^2 \beta_{-1}}{\epsilon\eta} \gamma_k \|u_k - u_{k-1}\|_{P_{k-1}} \|u_k - \hat{u}\|_{P_{k-1}}. \end{aligned}$$

At this point, we can apply Lemma 4 with $a_n = \|u_n - \hat{u}\|_{P_{n-1}}$, $b_n = \frac{\Lambda\|W\|^2}{\epsilon\eta} \beta_{-1} \|u_0 - \hat{u}\|_{P_{-1}}^2 + \frac{\Lambda\|W\|^2}{\epsilon\eta} \frac{\alpha_{-1}^2}{k_{\max}} \|v_0^0 - \hat{v}\|_{D_{-1}}^2 + \frac{\Lambda\|W\|^2}{\epsilon\eta} \sum_{k=0}^n \Lambda \beta_{-1} \gamma_k^2 \|u_k - u_{k-1}\|_{P_{k-1}}^2$ and $c_n = \frac{2\Lambda^2\|W\|^2 \beta_{-1}}{\epsilon\eta} \gamma_n \|u_n - u_{n-1}\|_{P_{n-1}}$, thus obtaining

$$\begin{aligned} \|u_{n+1} - \hat{u}\|_{P_n} & \leq \frac{1}{2} \sum_{k=0}^{n+1} \frac{2\Lambda^2\|W\|^2 \beta_{-1}}{\epsilon\eta} \gamma_k \|u_k - u_{k-1}\|_{P_{k-1}} \\ & + \left(\frac{\Lambda\|W\|^2}{\epsilon\eta} \left(\beta_{-1} \|u_0 - \hat{u}\|_{P_{-1}}^2 + \frac{\alpha_{-1}^2}{k_{\max}} \|v_0^0 - \hat{v}\|_{D_{-1}}^2 \right) \right. \\ & + \sum_{k=0}^n \Lambda \beta_{-1} \gamma_k^2 \|u_k - u_{k-1}\|_{P_{k-1}}^2 \left. \right)^{\frac{1}{2}} \\ & + \left(\frac{1}{2} \sum_{k=0}^{n+1} \frac{2\Lambda^2\|W\|^2 \beta_{-1}}{\epsilon\eta} \gamma_k \|u_k - u_{k-1}\|_{P_{k-1}} \right)^{\frac{1}{2}}. \end{aligned} \tag{40}$$

By recursively applying condition (32) to (31), we get

$$\begin{aligned} \sum_{k=0}^{\infty} \gamma_k \|u_k - u_{k-1}\|_{P_{k-1}} & \leq \sum_{k=0}^{\infty} \gamma_k \Lambda_k \|u_k - u_{k-1}\|_{P_{-1}} \\ & \leq \Lambda \|P_{-1}\|^{\frac{1}{2}} \sum_{k=0}^{\infty} \gamma_k \|u_k - u_{k-1}\| < \infty. \end{aligned} \tag{41}$$

By applying the previous inequality to (40) and recalling that $P_n \in \mathcal{D}_\eta$, it follows that the sequence $\{u_n\}_{n \in \mathbb{N}}$ is bounded. Finally, by discarding the term proportional to $\|u_{n+1} - \hat{u}\|_{P_n}^2$ in (39), employing again (41), the boundedness of $\{u_n\}_{n \in \mathbb{N}}$, the lower bound (25) on the eigenvalues of D_n and $\beta_{k-1} \leq \Lambda\beta_{-1}$, we conclude that also $\{v_n^0\}_{n \in \mathbb{N}}$ is bounded.

(ii) Item (i) guarantees the existence of a constant $M > 0$ such that $\|u_n - \hat{u}\|_{P_n} \leq M$ for all $n \geq 0$. Furthermore, we have $\beta_{n-1} \leq \Lambda\beta_{-1}$ for all $n \geq 0$. Then, from (38), we deduce the following inequality

$$\begin{aligned} & \beta_n k_{\max} \|u_{n+1} - \hat{u}\|_{P_n}^2 + \alpha_n^2 \|v_{n+1}^0 - \hat{v}\|_{D_n}^2 \\ & \leq (1 + \chi_{n-1})(\beta_{n-1} k_{\max} \|u_n - \hat{u}\|_{P_{n-1}}^2 + \alpha_{n-1}^2 \|v_n^0 - \hat{v}\|_{D_{n-1}}^2) \\ & \quad + \Lambda\beta_{-1} k_{\max} (1 + \chi_{n-1})(\gamma_n^2 \|u_n - u_{n-1}\|_{P_{n-1}}^2 + 2M\gamma_n \|u_n - u_{n-1}\|_{P_{n-1}}). \end{aligned}$$

The previous inequality and (41) allow us to apply Lemma 5 with $a_n = \beta_{n-1} k_{\max} \|u_n - \hat{u}\|_{P_{n-1}}^2 + \alpha_{n-1}^2 \|v_n^0 - \hat{v}\|_{D_{n-1}}^2$, $b_n = \chi_{n-1}$ and $c_n = \Lambda\beta_{-1} k_{\max} (1 + \chi_{n-1})(\gamma_n^2 \|u_n - u_{n-1}\|_{P_{n-1}}^2 + 2M\gamma_n \|u_n - u_{n-1}\|_{P_{n-1}})$, thus we get the thesis.

(iii) In virtue of (23)-(34), the sequences $\{\alpha_n\}_{n \in \mathbb{N}}$ and $\{\beta_n\}_{n \in \mathbb{N}}$ both satisfy the hypotheses of Lemma 5 and are bounded away from zero; hence, there exist $\alpha > 0$ and $\beta > 0$ such that

$$\lim_{n \rightarrow \infty} \alpha_n = \alpha, \quad \lim_{n \rightarrow \infty} \beta_n = \beta. \tag{42}$$

From inequality (41), we deduce that $\{P_n\}_{n \in \mathbb{N}} \subseteq \mathcal{D}_\eta^{\Lambda^2 \|P_{-1}\|}$; likewise, we have $\{D_n\}_{n \in \mathbb{N}} \subseteq \mathcal{D}_{1-\epsilon}^{\Lambda^2 \|D_{-1}\|}$. Thus, by Lemma 6, there exists $P \in \mathcal{D}_\eta$ and $D \in \mathcal{D}_{1-\epsilon}$ such that

$$\lim_{n \rightarrow \infty} P_n = P, \quad \lim_{n \rightarrow \infty} D_n = D. \tag{43}$$

Since $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ is bounded, there exists $(u^\dagger, v^\dagger) \in \mathbb{R}^d \times \mathbb{R}^{d'}$ and $I \subseteq \mathbb{N}$ such that

$$\lim_{n \in I} u_n = u^\dagger, \quad \lim_{n \in I} v_n^0 = v^\dagger. \tag{44}$$

Let us show that (u^\dagger, v^\dagger) is a solution of problem (12). To this aim, we consider inequality (30) with $u' = \hat{u}$, $v' = \hat{v}$, and by observing that $\mathcal{L}(u_{n+1}, \hat{v}) \geq \mathcal{L}(\hat{u}, \tilde{v}_n)$, applying relation $\bar{u}_n = u_n + \gamma(u_n - u_{n-1})$ and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \frac{1}{2} \left(\frac{1}{\alpha_n} - L_n \right) \|u_{n+1} - \bar{u}_n\|_{P_n}^2 + \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n^2}{2} \|v_n^k - v_n^{k+1}\|_{D_n}^2 \\ & \leq \frac{\beta_n}{2} (\|u_n - \hat{u}\|_{P_n}^2 - \|u_{n+1} - \hat{u}\|_{P_n}^2) + \frac{\alpha_n^2}{2k_{\max}} (\|v_n^0 - \hat{v}\|_{D_n}^2 - \|v_{n+1}^0 - \hat{v}\|_{D_n}^2) \\ & \quad + \frac{\beta_n \gamma_n^2}{2} \|u_n - u_{n-1}\|_{P_n}^2 + \beta_n \gamma_n \|u_n - \hat{u}\|_{P_{n-1}} \|u_n - u_{n-1}\|_{P_{n-1}}. \end{aligned} \tag{45}$$

Since $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ is bounded, there exist constants $M > 0$ and $\tilde{M} > 0$ such that

$$\|u_n - \hat{u}\|_{P_n} \leq M, \quad \|v_n^0 - \hat{v}\|_{D_n} \leq \tilde{M}, \quad \forall n \geq 0. \tag{46}$$

Additionally, by Step 7 of Algorithm 1, we observe that

$$\frac{1}{\alpha_n} - L_n = L_n \left(\frac{1}{\epsilon} - 1 \right) \geq L_0 \left(\frac{1}{\epsilon} - 1 \right) > 0, \quad \forall n \geq 0. \tag{47}$$

By employing bounds (46–47), conditions (32–35), the fact that $\alpha_n \leq \alpha_{n-1}$ and $\beta_n \leq (1 + \zeta_{n-1})\beta_{n-1} \leq (1 + \chi_{n-1})\beta_{n-1} \leq \Lambda\beta_{-1}$ as seen in (23) and (34), respectively, and the lower bounds on the eigenvalues of $\{P_n\}_{n \in \mathbb{N}}$ and $\{D_n\}_{n \in \mathbb{N}}$, we derive from (45) the following inequality

$$\begin{aligned} & \frac{\eta L_0}{2} \left(\frac{1}{\epsilon} - 1 \right) \|u_{n+1} - \bar{u}_n\|^2 + \frac{\min \left\{ \alpha_0^2, \frac{\epsilon^2 \delta^2 \eta^2}{L^2} \right\} (1 - \epsilon)}{2k_{\max}} \sum_{k=0}^{k_{\max}-1} \|v_n^k - v_n^{k+1}\|^2 \\ & \leq \frac{1}{2} (\beta_{n-1} \|u_n - \hat{u}\|_{P_{n-1}}^2 - \beta_n \|u_{n+1} - \hat{u}\|_{P_n}^2) \\ & + \frac{1}{2k_{\max}} (\alpha_{n-1}^2 \|v_n^0 - \hat{v}\|_{D_{n-1}}^2 - \alpha_n^2 \|v_{n+1}^0 - \hat{v}\|_{D_n}^2) + \frac{1}{2} \left(\Lambda\beta_{-1} M^2 + \frac{\alpha_{-1}^2 \tilde{M}^2}{k_{\max}} \right) \chi_{n-1} \\ & + \frac{\Lambda\beta_{-1} \gamma_n^2 (1 + \chi_{n-1})}{2} \|u_n - u_{n-1}\|_{P_{n-1}}^2 + \Lambda\beta_{-1} M \gamma_n \|u_n - u_{n-1}\|_{P_{n-1}}. \end{aligned}$$

We sum the previous relation over $n = 0, \dots, N$, thus obtaining

$$\begin{aligned} & \frac{\eta L_0}{2} \left(\frac{1}{\epsilon} - 1 \right) \sum_{n=0}^N \|u_{n+1} - \bar{u}_n\|^2 + \frac{\min \left\{ \alpha_0^2, \frac{\epsilon^2 \delta^2 \eta^2}{L^2} \right\} (1 - \epsilon)}{2k_{\max}} \sum_{n=0}^N \sum_{k=0}^{k_{\max}-1} \|v_n^k - v_n^{k+1}\|^2 \\ & \leq \frac{1}{2} (\beta_{-1} \|u_0 - \hat{u}\|_{P_{-1}}^2 - \beta_N \|u_{N+1} - \hat{u}\|_{P_N}^2) \\ & + \frac{1}{2k_{\max}} (\alpha_{-1}^2 \|v_0^0 - \hat{v}\|_{D_{-1}}^2 - \alpha_N^2 \|v_{N+1}^0 - \hat{v}\|_{D_N}^2) + \frac{1}{2} \left(\Lambda\beta_{-1} M^2 + \frac{\alpha_{-1}^2 \tilde{M}^2}{k_{\max}} \right) \sum_{n=0}^N \chi_{n-1} \\ & + \sum_{n=0}^N \frac{\Lambda\beta_{-1} \gamma_n^2 (1 + \chi_{n-1})}{2} \|u_n - u_{n-1}\|_{P_{n-1}}^2 + \sum_{n=0}^N \Lambda\beta_{-1} M \gamma_n \|u_n - u_{n-1}\|_{P_{n-1}}. \end{aligned}$$

Taking the limit for $N \rightarrow \infty$ and using conditions (35) and (41) yields

$$\sum_{n=0}^{\infty} \|u_{n+1} - \bar{u}_n\|^2 < \infty, \quad \sum_{n=0}^{\infty} \sum_{k=0}^{k_{\max}-1} \|v_n^k - v_n^{k+1}\|^2 < \infty,$$

which trivially leads to

$$\lim_{n \rightarrow \infty} \|u_{n+1} - \bar{u}_n\|^2 = 0, \quad \lim_{n \rightarrow \infty} \|v_n^k - v_n^{k+1}\| = 0, \quad k = 0, \dots, k_{\max} - 1. \tag{48}$$

Since $\lim_{n \in I} v_n^0 = v^\dagger$, the latter limits in (48) imply that

$$\lim_{n \in I} v_n^k = v^\dagger, \quad k = 1, \dots, k_{\max}. \tag{49}$$

Note also that, based on condition (31), we have

$$\lim_{n \rightarrow \infty} \|\bar{u}_n - u_n\| = \lim_{n \rightarrow \infty} \gamma_n \|u_n - u_{n-1}\| = 0,$$

which means that

$$\lim_{n \in I} \bar{u}_n = u^\dagger. \tag{50}$$

In turn, the above limit and the former one in (48) imply

$$\lim_{n \in I} u_{n+1} = u^\dagger. \tag{51}$$

By plugging limits (42), (43), (44), (49), (50) and (51) inside relation (A1) in Appendix A, we obtain

$$\nabla f(u^\dagger) + W^T v^\dagger = 0. \tag{52}$$

Likewise, by inserting the same limits inside relation (A3), we get

$$\lim_{n \in I} u_n^{k+1} = u^\dagger, \quad k = 0, \dots, k_{\max} - 1.$$

In particular, for $k = 0$, we have $\lim_{n \in I} u_n^1 = u^\dagger$. By combining this last fact with (42), (49) and the continuity of the operator $\text{prox}_{\beta\alpha^{-1}h^*}$ inside the dual step (21), we conclude that

$$v^\dagger = \text{prox}_{\beta\alpha^{-1}h^*}(v^\dagger + \beta\alpha^{-1}Wu^\dagger). \tag{53}$$

Thanks to equations (52) and (53), we can apply Lemma 7 and conclude that (u^\dagger, v^\dagger) is a solution of problem (12). Hence, it follows from item (ii) that the sequence $\{\beta_{n-1}k_{\max}\|u^\dagger - u_n\|_{P_{n-1}}^2 + \alpha_{n-1}^2\|v^\dagger - v_n^0\|_{D_{n-1}}^2\}_{n \in \mathbb{N}}$ converges and, by definition of limit point, it admits a subsequence converging to zero. Then the sequence $\{(u_n, v_n^0)\}_{n \in \mathbb{N}}$ converges to (u^\dagger, v^\dagger) . \square

4 Application to image deblurring

In this section, we focus our attention on the deblurring problem

$$\arg \min_{u \in \mathbb{R}^d} \frac{1}{2} \|Au - b\|_2^2 + \lambda TV(u), \tag{54}$$

where TV stands for the Total Variation operator, $\lambda > 0$ is the regularization parameter, $A \in \mathbb{R}^{d \times d}$ is the blurring operator and $b \in \mathbb{R}^d$ is the observed image affected by white Gaussian noise. Problem (54) can be reformulated as (1) by setting

$$f(u) = \frac{1}{2} \|Au - b\|_2^2, \quad h(Wu) = TV(u) = \sum_{i=1}^{d^2} \|\nabla_i u\|_2,$$

where $W = (\nabla_1^T, \dots, \nabla_{d^2}^T)^T \in \mathbb{R}^{2d^2 \times d}$ represents the discrete gradient operator and $h : \mathbb{R}^{2d^2} \rightarrow \mathbb{R} \cup \{\infty\}$ is defined as

$$h(t) = \sum_{i=1}^{d^2} \left\| \begin{pmatrix} t_{2i-1} \\ t_{2i} \end{pmatrix} \right\|.$$

Furthermore, it is straightforward to conclude that the minimization problem (54) satisfies Assumption 1.

4.1 Parameter choice

In the following, we detail how to appropriately select the parameters in Algorithm 1 applied to problem (54).

We equip Algorithm 1 with scaling matrices $\{P_n\}_{n \in \mathbb{N}}$ of the form

$$P_n = A^T A + v_n I_d, \quad \forall n \geq 0, \tag{55}$$

where $I_d \in \mathbb{R}^{d \times d}$ is the identity matrix and $\{v_n\}_{n \in \mathbb{N}}$ is either a positive constant sequence $v_n \equiv v > 0$ or a monotone sequence of positive real numbers converging to a positive value $v^* > 0$. These scaling matrices draw inspiration from the Iterated Tikhonov method [36–38]. Hence, from now on, we will refer to Algorithm 1 as NPDIT - Nested Primal–Dual Iterated Tikhonov method.

Note that the matrices $\{P_n\}_{n \in \mathbb{N}}$ defined in (55) can be diagonalized and inverted by two fast Fourier transforms (FFTs) since A is a convolution operator that can be diagonalized by FFT. Moreover, they comply with Assumption 2 (ii), as stated in the following result.

Lemma 11 *Let $\{v_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}_{>0}$ be a sequence that is either constant or monotone convergent to a positive value $v^* > 0$. Then, there exists a summable sequence of positive scalar $\{\zeta_n\}_{n \in \mathbb{N}}$ such that the sequence of matrices $\{P_n\}_{n \in \mathbb{N}}$ satisfies Assumption (2) (ii) that is*

$$P_n \preceq (1 + \zeta_{n-1})P_{n-1}, \quad \forall n \geq 0.$$

Proof If $\{v_n\}_{n \in \mathbb{N}}$ is a constant sequence such that $v_n \equiv v > 0$, then Assumption 2 (ii) trivially follows with $\zeta_n \equiv 0$.

If $\{v_n\}_{n \in \mathbb{N}}$ is a nonincreasing sequence converging to a positive value $v^* > 0$, then $P_n \in \mathcal{D}_{v_n}$ for all $n \geq 0$ and $P_n \leq P_{n-1}$ because $v_n \leq v_{n-1}$. Hence, the sequence $\{P_n\}_{n \in \mathbb{N}}$ satisfies Assumption 2 (ii) with $\zeta_n \equiv 0$.

Finally, if $\{v_n\}_{n \in \mathbb{N}}$ is a nondecreasing sequence converging to a positive value $v^* > 0$, then $P_n \in \mathcal{D}_{v_0}$ for all $n \geq 0$ and it follows that

$$\begin{aligned} u^T P_n u &= u^T (A^T A + v_n I_d) u \\ &= u^T (A^T A + v_{n-1} I_d) u + (v_n - v_{n-1}) u^T u \\ &\leq u^T P_{n-1} u + (v_n - v_{n-1}) v_0^{-1} u^T P_{n-1} u \\ &= (1 + (v_n - v_{n-1}) v_0^{-1}) u^T P_{n-1} u, \quad \forall u \in \mathbb{R}^d. \end{aligned}$$

where the inequality follows from the left-hand inequality in (6). By setting $\zeta_{n-1} = (v_n - v_{n-1}) v_0^{-1} \geq 0$, we note that

$$\sum_{n=1}^N \zeta_{n-1} = v_0^{-1} (v_N - v_0) \Rightarrow \sum_{n=1}^{\infty} \zeta_{n-1} < \infty.$$

Hence, the sequence $\{P_n\}_{n \in \mathbb{N}}$ satisfies Assumption 2 (ii). □

According to the selection of the parameter v_n in (55), we consider three different implementations of NPDIT. In the first one, we let

$$v_n \equiv v, \quad v > 0, \tag{56}$$

where $v > 0$ is a constant that must be prefixed in advance by the user. The selection of v significantly influences the performance and stability of the method throughout the iterations. Estimating it accurately can be challenging, which is why we have also explored two non-stationary strategies. More precisely, a non-stationary implementation - denominated NPDIT_D - adopts a decreasing geometric sequence for v_n defined as

$$v_n = a_0 q^n + v_{\text{final}}, \quad \forall n \geq 0 \tag{57}$$

where we set $a_0 = \frac{1}{2}$, $q = 0.85$, and $v_{\text{final}} = 10^{-2}$ in the numerical results. In a second non-stationary implementation - denominated NPDIT_I - employs an increasing sequence for v_n defined as

$$v_n = 1 - \frac{1}{n+1} + v_{\text{initial}}, \quad \forall n \geq 0 \tag{58}$$

with $v_{\text{initial}} = 10^{-2}$ in the numerical examples.

Initially, we also considered the trivial choice $P_n \equiv I_d$ for our tests; however, we noted that such a choice yields inferior performances when compared with the preconditioners (55). For this reason, we did not report the related results in the manuscript.

In the NPDIT algorithm and its two non-stationary variants, a backtracking strategy is employed to compute the steplength α_n . Particularly, at each step n , we compute $\alpha_n = \epsilon/L_n$, where $\epsilon = 0.99$ and $L_n > 0$ is such that

$$f(\tilde{u}_n) \leq f(\bar{u}_n) + \nabla f(\bar{u}_n)^T (\tilde{u}_n - \bar{u}_n) + \frac{L_n}{2} \|\tilde{u}_n - \bar{u}_n\|_{P_n}^2, \tag{59}$$

where \tilde{u}_n and \bar{u}_n are defined as in (22) and (19), respectively. We recall that the resulting sequence $\{\alpha_n\}_{n \in \mathbb{N}}$ is monotonically nonincreasing. We start with an initial value of $L_0 = 0.1$ and iteratively increase it by a factor of $\delta = 0.8$ until condition (59) is satisfied, or $L_n > 1$, in which case we simply set $L_n = 1$ and $\alpha_n = \epsilon$. This choice is justified by the following remark.

Remark 5 If $f(u) = \frac{1}{2}\|Au - b\|_2^2$ and $P_n = A^T A + \nu_n I$, then condition (59) is satisfied for $L_n = 1$. Indeed, using a second-order Taylor expansion at $u = \bar{u}_n$, we have

$$f(\tilde{u}_n) = f(\bar{u}_n) + \nabla f(\bar{u}_n)^T (\tilde{u}_n - \bar{u}_n) + \frac{1}{2}(\tilde{u}_n - \bar{u}_n)A^T A(\tilde{u}_n - \bar{u}_n)^T.$$

From the fact that the matrix $A^T A$ is positive semidefinite, we obtain

$$\begin{aligned} f(\tilde{u}_n) &\leq f(\bar{u}_n) + \nabla f(\bar{u}_n)^T (\tilde{u}_n - \bar{u}_n) + \frac{1}{2}(\tilde{u}_n - \bar{u}_n)(A^T A + \nu_n I)(\tilde{u}_n - \bar{u}_n)^T \\ &= f(\bar{u}_n) + \nabla f(\bar{u}_n)^T (\tilde{u}_n - \bar{u}_n) + \frac{1}{2}\|\tilde{u}_n - \bar{u}_n\|_{P_n}^2. \end{aligned}$$

The choice of the β_n parameter must be coherent with Step 3 of Algorithm 1. Therefore, in both nonstationary versions of the NPDIT method, we select β_n at each iteration $n \in \mathbb{N}$ as follows:

$$\beta_n = \frac{\epsilon \nu_n}{8} = \frac{\epsilon \nu_n}{\|W\|^2},$$

where $\epsilon \in (0, 1)$.

Finally, the value of γ_n for the inertial step was chosen following the strategy proposed in [34] and detailed below:

$$\gamma_n = \begin{cases} 0, & n = 0 \\ \min \left\{ \gamma_n^{\text{FISTA}}, \frac{C \rho_n}{\|u_n - u_{n-1}\|} \right\}, & n = 1, 2, \dots \end{cases} \tag{60}$$

In all of the three considered examples, we chose $C = 0.1\|u_1 - u_0\|$ for the stationary version of NPDIT and for NPDIT_D, while we selected $C = \|u_1 - u_0\|$ for the increasing case. Then, for all NPDIT implementations, we used the same value for ρ_n , which is

$\rho_n = \frac{1}{n^{1.1}}$. In (60), the parameter γ_n^{FISTA} is defined by the usual FISTA rule:

$$t_0 = 1, \quad \begin{cases} t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}, \\ \gamma_n^{\text{FISTA}} = \frac{t_{n-1}}{t_{n+1}}, \end{cases} \quad n = 0, 1, \dots \quad (61)$$

The number of primal–dual iterations is set for simplicity to $k_{\max} = 1$. This choice is coherent with the numerical experimentation carried out in [34] for a special instance of Algorithm 1, where it was observed that $k_{\max} = 1$ could be seen as a good trade-off between accuracy and complexity.

Finally, we remark that an automated method for computing the regularization parameter λ has not been implemented in the tests. Consequently, we will specify the value or values used for each example.

4.2 Numerical results

The following numerical results were obtained using MATLAB R2023a on a MacBook Pro equipped with the M2 Pro chip and a 10-core CPU (6 performance cores and 4 efficiency cores), along with 16GB of RAM.

We provide three image deblurring examples in which we compare the three implementations of our NPDIT method with the following competitors:

- The Nested Primal–Dual (NPD) method proposed in [34], which can be seen as a special instance of Algorithm 1 obtained by setting $P_n \equiv I_d$, $\eta_n \equiv 1$, $\alpha_0 = 1$, $\beta_n \equiv \beta < 1/8$ and γ_n computed according to (60) with $C = 10\|u_1 - u_0\|$; unlike Algorithm 1, NPD chooses a fixed steplength $\alpha_n \equiv \alpha_0$ without performing any backtracking procedure; note that the chosen value for α_0 is such that the backtracking condition at Step 7 of Algorithm 1 is automatically satisfied for all $n \geq 0$ with $\alpha_n = \alpha_0$;
- The preconditioned version of the Chambolle–Pock method (CP_{prec}), the popular primal–dual algorithm for composite convex optimization proposed in [44];
- A first order primal–dual method with linesearch (PD–LS) proposed by Malitsky and Pock in [17];
- The inexact version of FISTA with approximate proximal evaluations proposed in [29], whose practical implementation resembles Algorithm 1, in the sense that it can be seen as a nested primal–dual algorithm with extrapolation for solving problems of the form (1); however, unlike Algorithm 1, the number of inner iterations varies with the outer iteration according to an appropriate stopping criterion, the extrapolation parameter is computed as proposed by Chambolle and Dossal in [45], and there is no variable metric.

The blurred image is obtained by circular two-dimensional convolution, ensuring that the deblurring model is not affected by boundary ringing effects. Consequently, the numerical results only depend on the applied numerical method. Of course, in real applications, appropriate boundary conditions affecting the structure of the matrix A should be adopted, see [46].

In order to evaluate the performance of the different methods, we use the Relative Reconstruction Error (RRE) functional, defined as

$$\text{RRE}(u_n) = \frac{\|u_{\text{true}} - u_n\|}{\|u_{\text{true}}\|},$$

where u_{true} represents the original image.

The relative decrease of the objective function is computed by the quantity

$$\frac{R(u_n) - R(u^*)}{R(u^*)},$$

where R is the objective function and u^* is the best reconstruction obtained among all the methods considered, achieving the minimum value of the objective function. This solution was precomputed by running all the methods for 10^3 iterations.

Example 1: Cameraman with 1% of noise.

In this example, we considered an image of a Cameraman with dimensions of 256×256 . The image was blurred using a 10×10 Gaussian Point Spread Function (PSF) with a standard deviation of 2. Additionally, white noise with an intensity level of 1% was added to the blurred image. This implies that the norm of the noise is equal to 1% of the norm of the blurred image. For this example, we set $\lambda = 10^{-4}$ as the regularization parameter.

In the first row of Fig. 1, we depict the behavior of the RRE functional over 400 iterations (left) and over 2 s of time (right) for all the methods considered. For the stationary NPDIT case, we set $\nu = 0.01$. We observe that our proposal, both stationary and non-stationary, outperforms all the other competitors. Indeed, we achieve lower values of the RRE in fewer iterations, which are also faster to compute compared to all the other cases. For NPDIT₁, i.e., the version of our proposal where ν_n is increasing, we note a significant speed-up in the initial iterations due to the small value of ν_n . However, as iterations progress, the performance worsens because ν_n approaches 1 and consequently slows down the convergence of the method.

In the second row of Fig. 1, we compare the relative decrease of the objective function. Similarly, the stationary method and the non-stationary method NPDIT-D achieve better results than all the other competitors. However, there are two observations worth mentioning. The inexact version of FISTA proposed in [29] seems to perform comparably to NPDIT-D as the iterations progress, but the time required to achieve these results is significantly greater than that required by our proposal. Secondly, the performance of NPDIT₁ is similar to the standard NPD algorithm. This is mainly due to the choice of the regularization parameter λ . As we will show in the last part of the numerical examples, NPDIT₁ works better when the regularization parameter λ is underestimated.

Figure 2 shows the convergence behavior of the RRE functional for NPDIT₁, NPDIT_D, and the stationary variant of NPDIT equipped with three different values of the parameter ν . The plots for the non-stationary methods and the case with $\nu = 0.01$ are identical to those shown in Fig. 1. For $\nu = 10^{-3}$, it is noticeable that the method introduces some instability and the speed of convergence is also reduced.

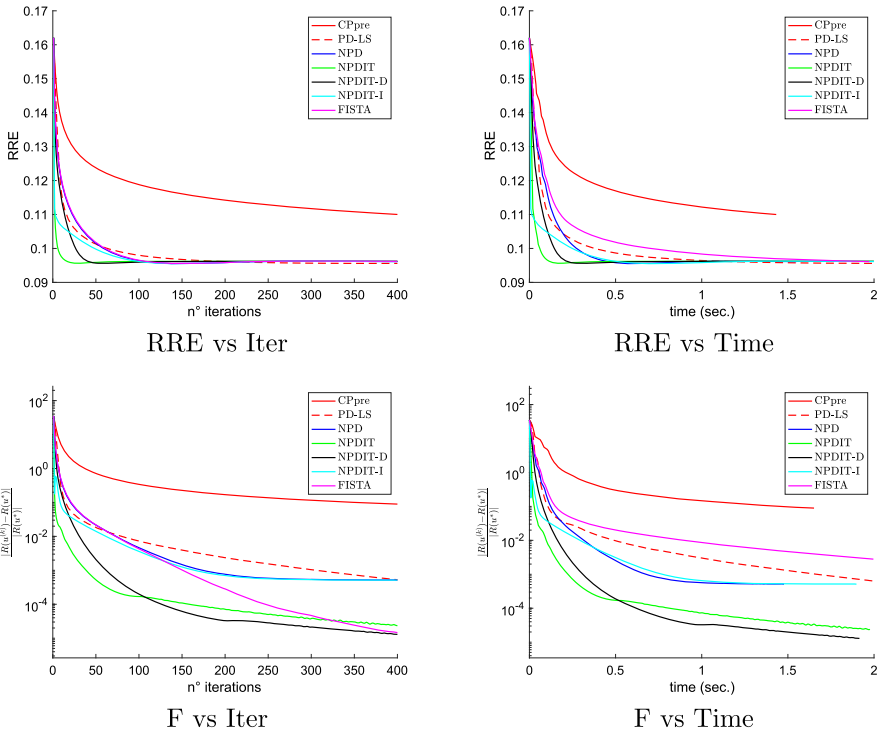
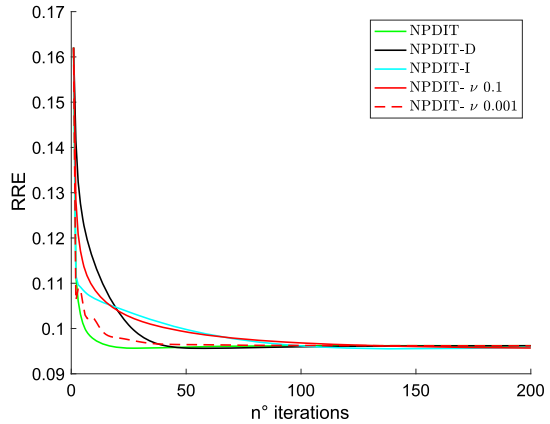


Fig. 1 Example 1—First row: RRE functional over 400 iterations and over 2 s of time. Second row: Relative decrease of the objective function over the number of iterations and time

Fig. 2 Example 1—RRE for different values of ν for $\lambda = 10^{-4}$



When $\nu = 0.1$, the algorithm is stable again but it is slower than in the other cases. This demonstrates that the convergence speed and the stability of the method are affected by the choice of ν for stationary values $\nu_n = \nu$. Therefore, the decreasing sequence of ν_n in equation (57) provides a good trade-off without requiring the estimation of ν .

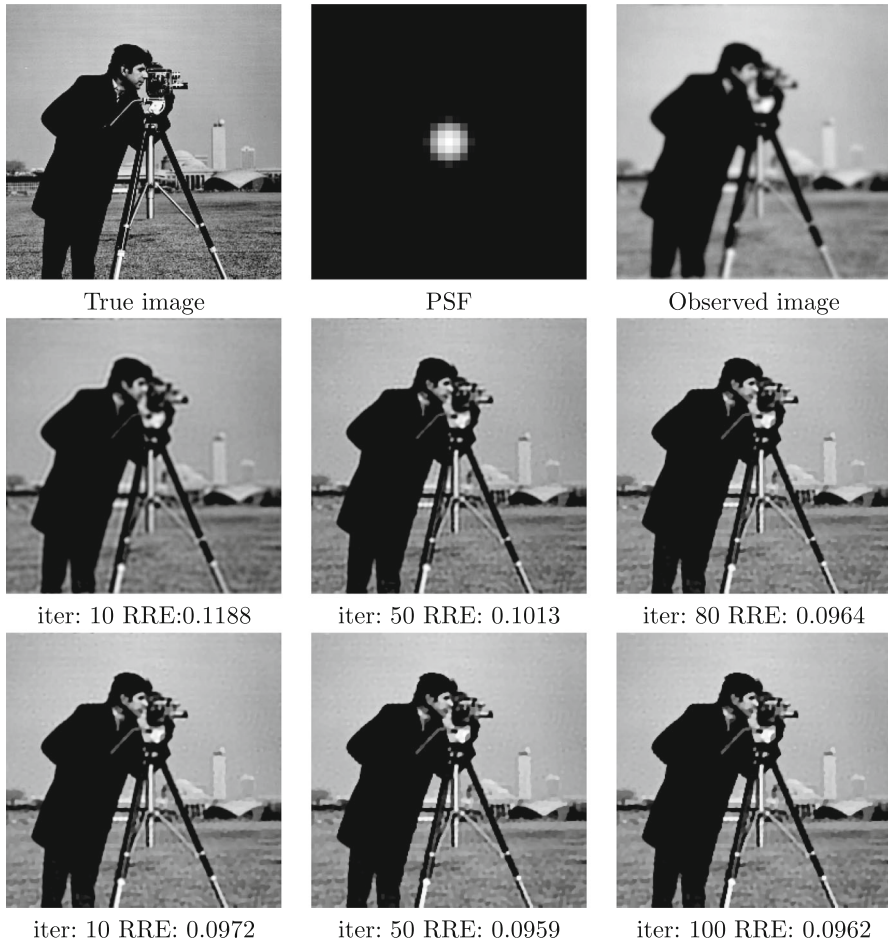


Fig. 3 Example 1—Reconstruction obtained at different iterations. First row: original data. Second row: images obtained with NPD. Third row: images obtained with NPDIT ($\nu = 0.01$)

The reconstructed images shown in Fig. 3 are obtained at various iterations for both the NPD and the NPDIT method with $\nu = 0.01$. For completeness, the upper part of the figure includes the true image, the PSF, and the observed image b . Additionally, the achieved RRE value is reported for each reconstruction. Notably, the NPDIT method achieves satisfactory reconstructions with just 20 iterations, while the NPD method requires at least 100 iterations to achieve a similar level of accuracy.

Example 2: Peppers.

In this second example, we examined the image of Peppers with dimensions 256×256 and applied an out-of-focus PSF with dimensions 10×10 for blurring. Additionally, we introduced white Gaussian noise with an intensity level of 1% to the image. For this case, we chose $\lambda = 10^{-4}$ as the regularization parameter.

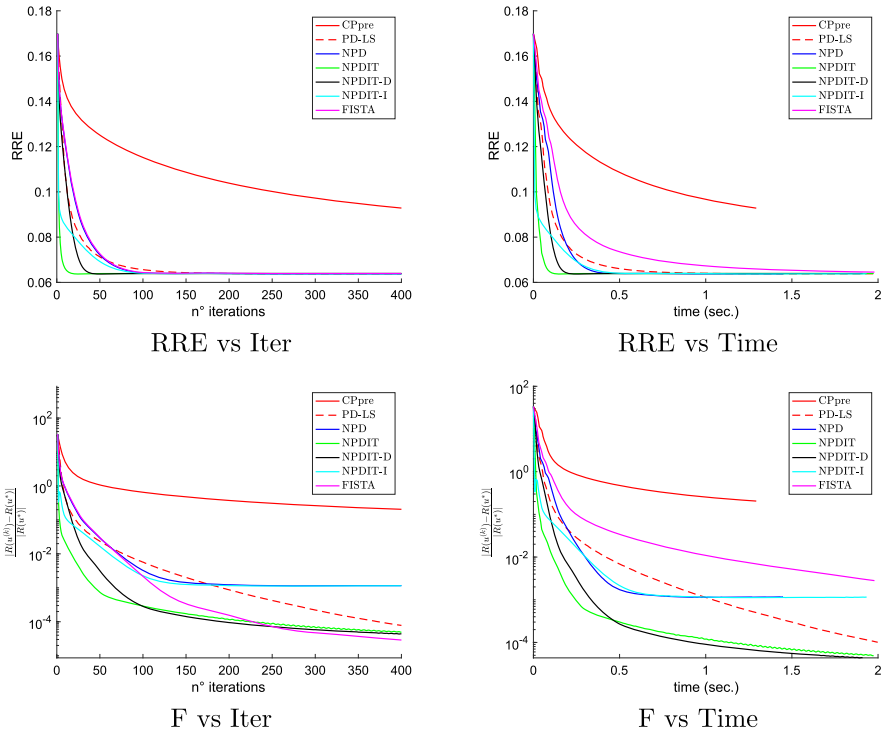


Fig. 4 Example 2—First row: RRE functional over 400 iterations and over 2 s of time. Second row: Relative decrease of the objective function over the number of iterations and time

The comparison between all the methods follows the same structure as in the previous case. In the first row of Fig. 4, we present the RRE functional, while in the second row we focus on the relative decrease of the objective function. In this example, we observe that both the NPDIT method with $\nu = 0.01$ and the non-stationary version $NPDIT_D$ achieve excellent results. Concerning the $NPDIT_I$ algorithm, we notice a rapid decrease in the first iterations for the objective function, followed by a deterioration in performance as the iterations increase. Again, we note that the FISTA method outperforms all of our proposals in the relative decrease of the objective function after 400 iterations. However, as in Example 1, the time required is significantly greater than those of the stationary NPDIT and the $NPDIT_D$ method.

Finally, in Fig. 5, we display the reconstructions obtained by the NPD method and the NPDIT with $\nu = 0.01$ fixed. Once again, in the first row, we present the true image, the PSF used for blurring, and the observation corrupted by noise. For this example, it is evident that the stationary version NPDIT requires only 10 iterations to achieve excellent results, while the NPD algorithm needs almost four times more iterations to obtain the same quality in the reconstruction.

Example 3: Overlapping group TV

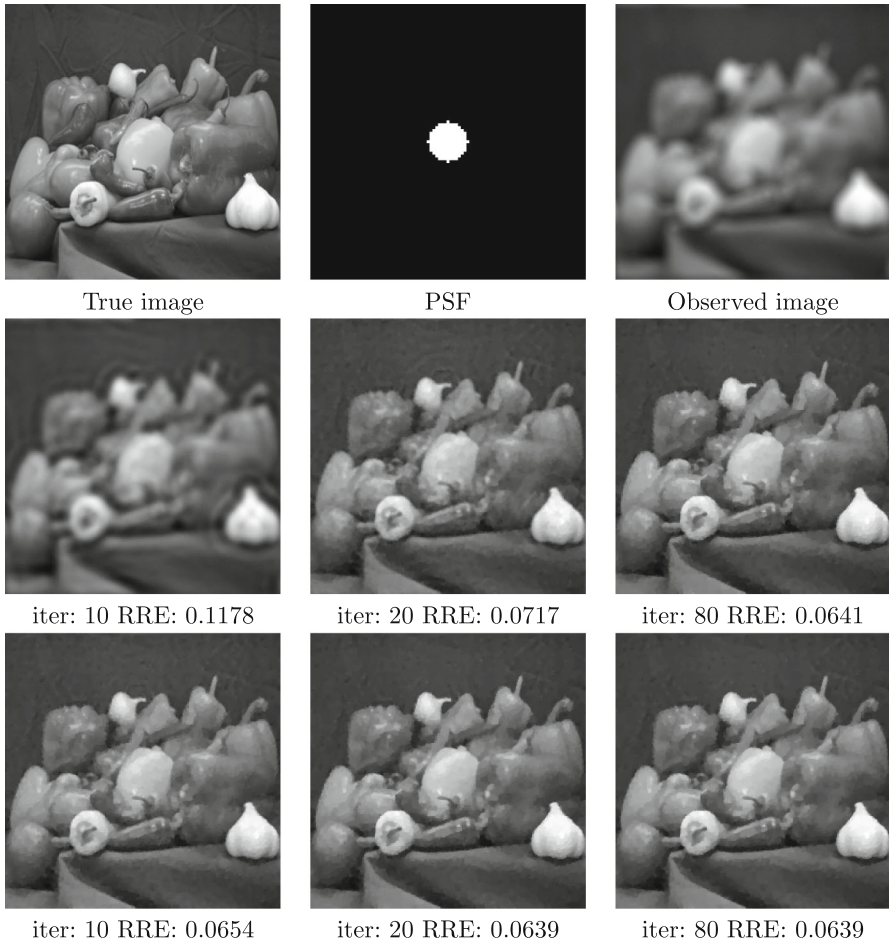


Fig. 5 Example 2—Reconstruction obtained at different iterations. First row: original data. Second row: images obtained with NPD. Third row: images obtained with NPDIT ($\nu = 0.01$)

In this final numerical example, we replace the TV term with the Overlapping Group sparsity TV (OG-TV) regularizer proposed in [47], which is defined as

$$h(Wx) = \sum_{i=1}^{d^2} \|B_i \nabla^h x\| + \|B_i \nabla^v x\|,$$

where $\nabla^h, \nabla^v \in \mathbb{R}^{d^2 \times d^2}$ are the horizontal and vertical discrete gradient operators, while $B_i \in \mathbb{R}^{m^2 \times d^2}$ is a unitary diagonal matrix such that $B_i x$ is the vector whose entries are the pixel i of the image x and its $m^2 - 1$ first neighbors. In this case we have that $\|W\|^2 = 8m^2$.

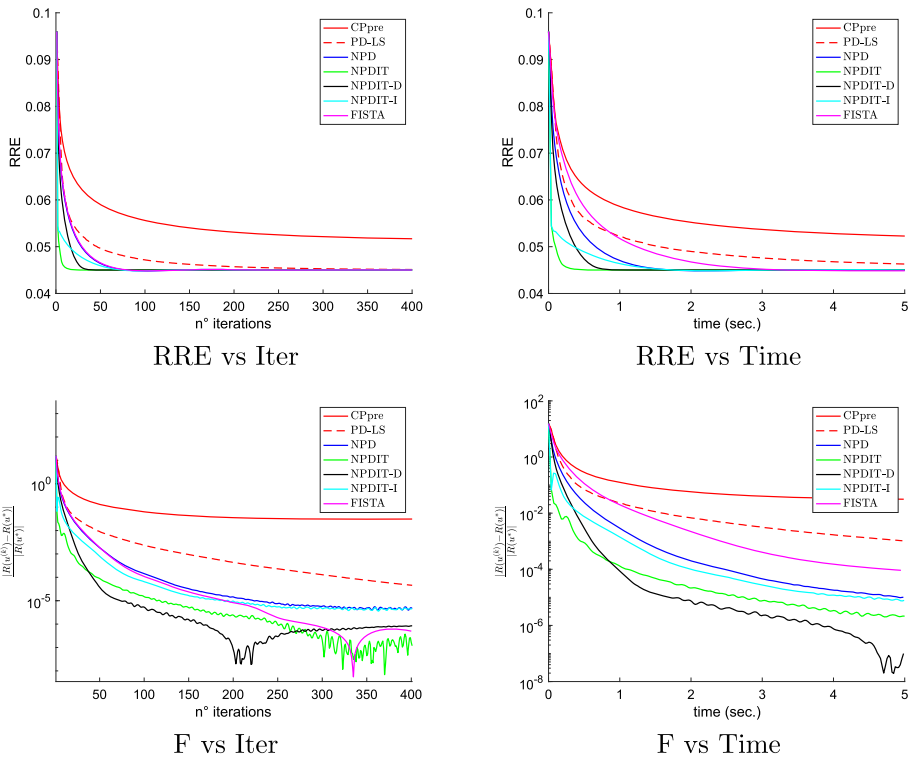


Fig. 6 Example 3—First row: RRE functional over 400 iterations and over 5 s of time. Second row: Relative decrease of the objective function over the same number of iterations and time as before

In order to test the performances of our methods on this new problem, we consider the jetplane image, which has been blurred and artificially corrupted with white Gaussian noise as in Example 1. For this case, we choose $\lambda = 5 \times 10^{-5}$ as the regularization parameter.

The analysis of this experiment follows the same structure as in the previous examples. In the first row of Fig. 6, we report the RRE functional as a function of the number of iterations (left side) and time (right side). The second row is devoted to the relative decrease of the objective function, again depending on the number of iterations and time. Even in this case, we can clearly state that all our proposals outperform the other methods in terms of both quality and speed.

Lastly, in Fig. 7, we compare the reconstruction obtained with the NPD method and the stationary NPDIT method equipped with $\nu = 0.01$. The final images were computed after different numbers of iterations, and we can see that, even in this case, NPDIT achieves remarkable reconstructions even after just 10 iterations.

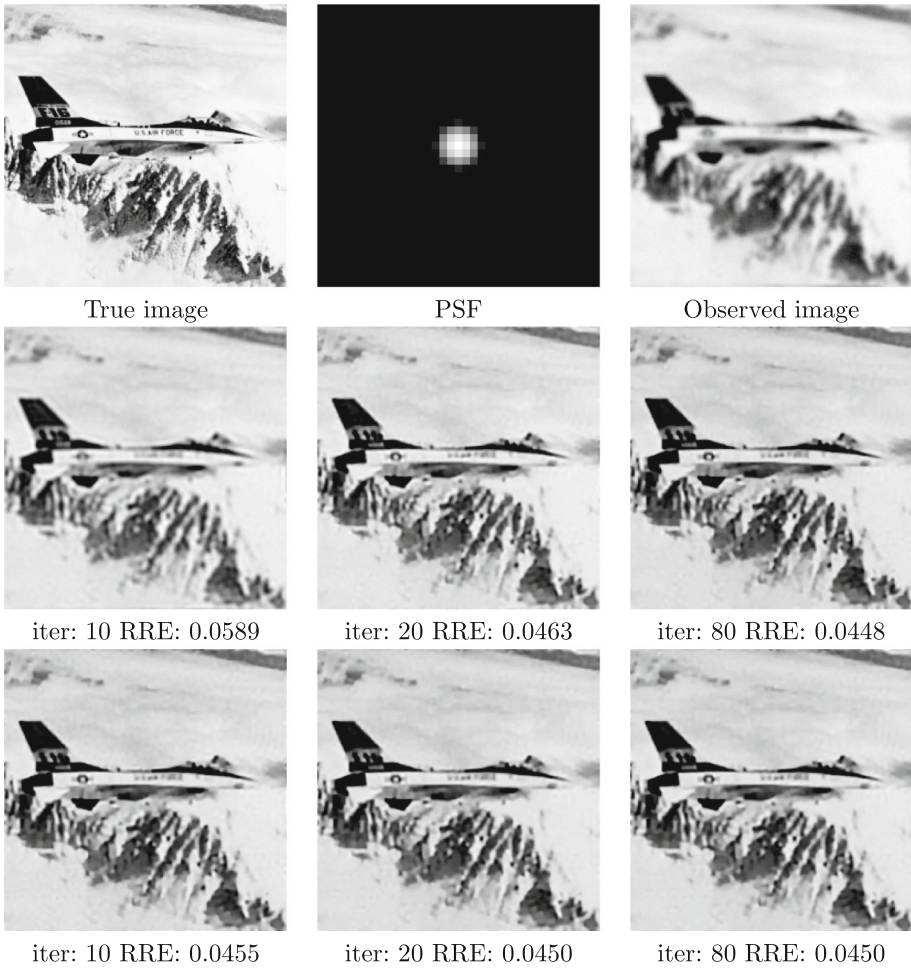


Fig. 7 Example 3—Reconstruction obtained at different iterations. First row: original data. Second row: images obtained with NPD. Third row: images obtained with NPDIT ($\nu = 0.01$)

4.3 Stability

As a final illustration, we investigated the robustness of the methods with respect to the regularization parameter λ under the condition of a stationary sequence $\nu_n = \nu$. To achieve this, we explored different values of this parameter for each example presented earlier. Furthermore, we conducted a detailed analysis of the behavior of the proposed method with respect to the two non-stationary sequences employed in NPDIT_D and NPDIT_I.

The comparison among the methods is carried out by assessing the quality of reconstructions using the RRE functional. Figure 8 presents a summary of the results as follows: each row corresponds to one of the mentioned examples. In particular, the i -th row corresponds to Example i , where $i = 1, 2, 3$. Each column is associated

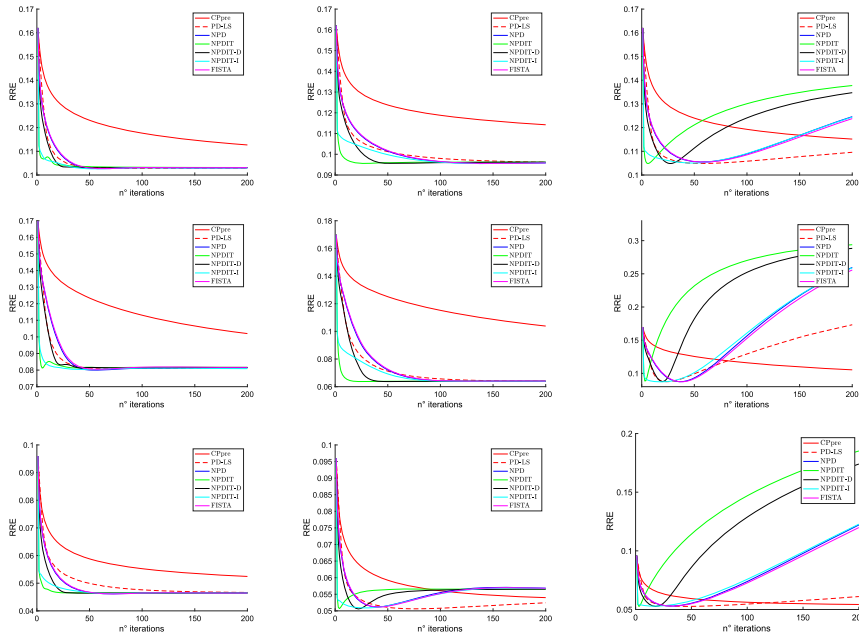


Fig. 8 From top to bottom: RRE functional with respect to the number of iterations for example 1 (top row), example 2 (middle row) and example 3 (bottom row). From left to right: comparison for different values of λ , i.e., $\lambda \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ for the first two rows, $\lambda \in \{10^{-4}, 10^{-5}, 10^{-6}\}$ for the third one

with a specific value of $\lambda \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ for the first two rows and $\lambda \in \{10^{-4}, 10^{-5}, 10^{-6}\}$ for the last one.

In accordance with the analysis conducted in Sect. 4.2, the NPDIT_D method exhibits behavior very similar to that of NPDIT , avoiding the selection of the stationary parameter ν . Regarding the NPDIT_I method, it is observed to be more beneficial in cases of underestimation of the regularization parameter λ , as indicated by the theoretical analysis in [48] for the Iterated Tikhonov method. Specifically, from the rightmost column, we observe greater stability around the minimum point compared to other methods. However, whenever λ is underestimated, it is advisable to implement an effective stopping criterion (e.g., the discrepancy principle) to terminate the process before the RRE grows excessively.

5 Conclusions and future work

In this paper, we have formulated and analysed a nested primal–dual method tailored for solving regularized convex optimization problems. Our proposed approach approximates a variable metric proximal–gradient step with extrapolation by executing a fixed number of primal–dual iterates, while dynamically adjusting the steplength parameter through a specialized backtracking procedure. The convergence of the iterates sequence towards a solution of the problem has been established by assuming a

relaxed monotonicity condition on the scaling matrices, as well as a specific shrinking criterion on the extrapolation parameters. Extensive numerical experiments have shown that our algorithm performs well in comparison to other similar competitors on some Total-Variation regularized least-squares problems arising in image deblurring, especially when it is equipped with scaling matrices inspired by the Iterated Tikhonov method.

Future work may include the application of our algorithm to other imaging problems such as computer tomography, its adaptation to more accurate boundary conditions for image deblurring, and its extension to more general optimization problems. In particular, for applications not addressed in this paper, the matrix–vector product with the proposed scaling matrices might be computationally expensive. Therefore, we plan to investigate its approximation in suitable subspaces from both a theoretical and practical standpoint. Moreover, we will investigate the extension of our proposed algorithm to problems where constraints on the primal variable are imposed, and the definition of novel nested primal–dual algorithms where the backtracking procedure based on the Descent Lemma is replaced by a linesearch along the descent direction.

Appendix A Proof of Lemma 10

Proof (i) We start by observing that

$$\begin{aligned}
 u_{n+1} = \tilde{u}_n &= \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} u_n^{k+1} \\
 &= \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} (\bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n) - \alpha_n P_n^{-1} W^T v_n^{k+1}) \\
 &= \bar{u}_n - \alpha_n P_n^{-1} \nabla f(\bar{u}_n) - \alpha_n P_n^{-1} W^T \tilde{v}_n.
 \end{aligned}
 \tag{A1}$$

Then, we can proceed from the value $\mathcal{L}(u', \tilde{v}_n)$ as follows:

$$\begin{aligned}
 \mathcal{L}(u', \tilde{v}_n) &= f(u') + \langle W u', \tilde{v}_n \rangle - h^*(\tilde{v}_n) \\
 &\geq f(\bar{u}_n) + \langle u' - \bar{u}_n, \nabla f(\bar{u}_n) \rangle + \langle W u', \tilde{v}_n \rangle - h^*(\tilde{v}_n) \\
 &= f(\bar{u}_n) + \langle u_{n+1} - \bar{u}_n, \nabla f(\bar{u}_n) \rangle - \langle u_{n+1} - u', \nabla f(\bar{u}_n) \rangle \\
 &\quad + \langle W u', \tilde{v}_n \rangle - h^*(\tilde{v}_n) \\
 &\geq f(u_{n+1}) - \frac{L_n}{2} \|u_{n+1} - \bar{u}_n\|_{P_n}^2 - \langle u_{n+1} - u', \nabla f(\bar{u}_n) \rangle \\
 &\quad + \langle W u', \tilde{v}_n \rangle - h^*(\tilde{v}_n),
 \end{aligned}$$

where the first inequality follows from the convexity of f , and the second one is based on the backtracking condition at Step 7 of Algorithm 1. By summing and subtracting the term $W^T \tilde{v}_n$ in the second argument of the scalar product $\langle u_{n+1} - u', \nabla f(\bar{u}_n) \rangle$, we

can extend the chain of inequalities as below:

$$\begin{aligned}
 \mathcal{L}(u', \tilde{v}_n) &\geq f(u_{n+1}) - \frac{L_n}{2} \|u_{n+1} - \bar{u}_n\|_{P_n}^2 - \langle u_{n+1} - u', \nabla f(\bar{u}_n) + W^T \tilde{v}_n \rangle \\
 &\quad + \langle Wu', \tilde{v}_n \rangle + \langle u_{n+1} - u', W^T \tilde{v}_n \rangle - h^*(\tilde{v}_n) \\
 &= f(u_{n+1}) + \langle u_{n+1}, W^T \tilde{v}_n \rangle - h^*(\tilde{v}_n) \\
 &\quad - \frac{1}{\alpha_n} \langle u_{n+1} - u', P_n(\alpha_n P_n^{-1} \nabla f(\bar{u}_n) + \alpha_n P_n^{-1} W^T \tilde{v}_n) \rangle \\
 &\quad - \frac{L_n}{2} \|u_{n+1} - \bar{u}_n\|_{P_n}^2 \\
 &= \mathcal{L}(u_{n+1}, \tilde{v}_n) + \frac{1}{\alpha_n} \langle u_{n+1} - u', P_n(u_{n+1} - \bar{u}_n) \rangle - \frac{L_n}{2} \|u_{n+1} - \bar{u}_n\|_{P_n}^2 \\
 &= \mathcal{L}(u_{n+1}, \tilde{v}_n) + \frac{1}{2\alpha_n} \|u_{n+1} - u'\|_{P_n}^2 + \frac{1}{2\alpha_n} \|u_{n+1} - \bar{u}_n\|_{P_n}^2 \\
 &\quad - \frac{1}{2\alpha_n} \|\bar{u}_n - u'\|_{P_n}^2 - \frac{L_n}{2} \|u_{n+1} - \bar{u}_n\|_{P_n}^2,
 \end{aligned}$$

where the second equality follows from (A1) together with the definition of \mathcal{L} in (12), and the third one is due to the application of the three-point equality

$$\|a - c\|_{P_n}^2 = \|a - b\|_{P_n}^2 + \|b - c\|_{P_n}^2 + 2\langle a - b, P_n(b - c) \rangle,$$

with $a = u', b = u_{n+1}, c = \bar{u}_n$. By rearranging the various terms, we get (28). (ii)

Starting from the value $\mathcal{L}(u_{n+1}, v_n^{k+1})$ and applying Lemma 3 with $\varphi = \beta_n \alpha_n^{-1} h^*$, $y = v_n^{k+1}, x = v_n^k, e = \beta_n \alpha_n^{-1} W u_n^k$ and $z = v'$, we get

$$\begin{aligned}
 \mathcal{L}(u_{n+1}, v_n^{k+1}) &= f(u_{n+1}) + \langle W u_{n+1}, v_n^{k+1} \rangle - h^*(v_n^{k+1}) \\
 &\geq f(u_{n+1}) + \langle W u_{n+1}, v_n^{k+1} \rangle - h^*(v') - \langle v_n^{k+1} - v', W u_n^k \rangle \\
 &\quad + \frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|^2 + \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|^2 \\
 &= \mathcal{L}(u_{n+1}, v') + \langle v_n^{k+1} - v', W(u_{n+1} - u_n^{k+1}) \rangle \\
 &\quad - \langle v_n^{k+1} - v', W(u_n^k - u_n^{k+1}) \rangle \\
 &\quad + \frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|^2 + \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|^2. \tag{A2}
 \end{aligned}$$

From Step 4 of Algorithm 1 and (A1), we also have the following relations

$$\begin{aligned}
 u_{n+1} - u_n^{k+1} &= \alpha_n P_n^{-1} W^T (v_n^{k+1} - \tilde{v}_n) \\
 u_n^k - u_n^{k+1} &= \alpha_n P_n^{-1} W^T (v_n^{k+1} - v_n^k). \tag{A3}
 \end{aligned}$$

Plugging the above relations inside (A2) yields

$$\mathcal{L}(u_{n+1}, v_n^{k+1}) \geq \mathcal{L}(u_{n+1}, v') + \alpha_n \langle W^T (v_n^{k+1} - v'), P_n^{-1} W^T (v_n^{k+1} - \tilde{v}_n) \rangle$$

$$\begin{aligned}
 & -\alpha_n \langle W^T(v_n^{k+1} - v'), P_n^{-1}W^T(v_n^{k+1} - v_n^k) \rangle \\
 & + \frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|^2 + \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|^2. \tag{A4}
 \end{aligned}$$

Note that the first scalar product in the previous inequality can be lower bounded as follows

$$\begin{aligned}
 \langle W^T(v_n^{k+1} - v'), P_n^{-1}W^T(v_n^{k+1} - \tilde{v}_n) \rangle &= \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(v_n^{k+1} - v')\|^2 \\
 &+ \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(v_n^{k+1} - \tilde{v}_n)\|^2 \\
 &- \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(\tilde{v}_n - v')\|^2 \\
 &\geq \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(v_n^{k+1} - v')\|^2 \\
 &- \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(\tilde{v}_n - v')\|^2,
 \end{aligned}$$

whereas the second one can be rewritten as

$$\begin{aligned}
 -\langle W^T(v_n^{k+1} - v'), P_n^{-1}W^T(v_n^{k+1} - v_n^k) \rangle &= -\frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(v_n^{k+1} - v')\|^2 \\
 &- \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(v_n^{k+1} - v_n^k)\|^2 \\
 &+ \frac{1}{2} \|P_n^{-\frac{1}{2}}W^T(v_n^k - v')\|^2.
 \end{aligned}$$

By inserting the previous relations inside (A4) and recalling the definition of D_n and $\|\cdot\|_{D_n}$ in (24–26), we come to

$$\begin{aligned}
 \mathcal{L}(u_{n+1}, v_n^{k+1}) &\geq \mathcal{L}(u_{n+1}, v') - \frac{\alpha_n}{2} \|P_n^{-\frac{1}{2}}W^T(\tilde{v}_n - v')\|^2 \\
 &+ \frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|_{D_n}^2 + \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2. \tag{A5}
 \end{aligned}$$

We proceed as follows

$$\begin{aligned}
 \mathcal{L}(u_{n+1}, \tilde{v}_n) &= \mathcal{L}\left(u_{n+1}, \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} v_n^{k+1}\right) \\
 &\geq \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \mathcal{L}(u_{n+1}, v_n^{k+1}) \\
 &\geq \mathcal{L}(u_{n+1}, v') - \frac{\alpha_n}{2} \|P_n^{-\frac{1}{2}}W^T(\tilde{v}_n - v')\|^2 \\
 &+ \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \left[\frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|_{D_n}^2 + \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2 \right]
 \end{aligned}$$

$$\begin{aligned} &\geq \mathcal{L}(u_{n+1}, v') - \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2} \|P_n^{-\frac{1}{2}} W^T (v_n^{k+1} - v')\|^2 \\ &+ \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|_{D_n}^2 + \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2, \end{aligned}$$

where the first inequality is obtained by concavity of $\mathcal{L}(u_{n+1}, \cdot)$, the second one from applying (A5), and the third one follows from the definition of \tilde{v}_n and the concavity of $-\|P_n^{-\frac{1}{2}} W^T (\cdot - v')\|^2$. We further extend the above chain of inequalities as done below

$$\begin{aligned} \mathcal{L}(u_{n+1}, \tilde{v}_n) &\geq \mathcal{L}(u_{n+1}, v') + \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^{k+1} - v'\|_{D_n}^2 - \frac{\alpha_n}{2\beta_n} \|v_n^k - v'\|_{D_n}^2 \\ &+ \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2 \\ &= \mathcal{L}(u_{n+1}, v') + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_n^{k_{\max}} - v'\|_{D_n}^2 - \frac{\alpha_n}{2\beta_n k_{\max}} \|v_n^0 - v'\|_{D_n}^2 \\ &+ \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2 \\ &= \mathcal{L}(u_{n+1}, v') + \frac{\alpha_n}{2\beta_n k_{\max}} \|v_{n+1}^0 - v'\|_{D_n}^2 - \frac{\alpha_n}{2\beta_n k_{\max}} \|v_n^0 - v'\|_{D_n}^2 \\ &+ \frac{1}{k_{\max}} \sum_{k=0}^{k_{\max}-1} \frac{\alpha_n}{2\beta_n} \|v_n^k - v_n^{k+1}\|_{D_n}^2, \end{aligned}$$

where the first inequality has been obtained by recalling once again the definition of D_n in (24), and the last equality is a consequence of the warm-start strategy at Step 3 of Algorithm 1. By rearranging terms, we finally get inequality (29).

(iii) Inequality (30) follows by summing (28) with (29). \square

Acknowledgements The authors are all members of the INdAM research group GNCS, which is kindly acknowledged.

Author Contributions All authors contributed to the study conception and design of the paper. Most of the theoretical results have been proved by S. Rebegoldi. Code implementation, data collection and analysis were performed by S. Aleotti. The first draft of the manuscript was written by S. Aleotti and S. Rebegoldi, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Modena e Reggio Emilia within the CRUI-CARE Agreement. S. Bonettini, M. Donatelli and S. Aleotti are supported by the Italian MUR through the PRIN 2022 project “Inverse Problems in the Imaging Sciences (IPIS)”, project code: 2022ANC8HL (CUP E53D23005580006), under the National Recovery and Resilience Plan (PNRR), Italy, Mission 04 Component 2 Investment 1.1 funded by the European Commission - NextGeneration EU programme. M. Prato is supported by the Italian MUR through the PRIN 2022 project “STILE: Sustainable Tomographic Imaging with Learning and rEgularization”, project code: 20225STXSB (CUP E53D23005480006), under the National Recovery and Resilience Plan (PNRR), Italy, Mission 04 Component 2 Investment 1.1 funded

by the European Commission - NextGeneration EU programme. S. Rebegoldi is supported by the Italian MUR through the PRIN 2022 project “Inverse problems in PDE: theoretical and numerical analysis”, project code: 2022B32J5C (CUP B53D23009200006), and the PRIN 2022 PNRR Project “Advanced optimization METHODS for automated central veIn Sign detection in multiple sclerosis from magneTic resonAnce imaging (AMETISTA)”, project code: P2022J9SNP (CUP E53D23017980001), under the National Recovery and Resilience Plan (PNRR), Italy, Mission 04 Component 2 Investment 1.1 funded by the European Commission—NextGeneration EU programme.

Data availability The test images *cameraman* and *peppers* in the experiments described in Sect. 4.2 are included in the Image Processing Toolbox of Matlab.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethical approval The authors declare that research ethics approval was not required for this study.

Informed consent The authors declare that informed consents were not required for this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bach, F., Jenatton, R., Mairal, J., Obozinski, G.: Structured sparsity through convex optimization. *Stat. Sci.* **27**(4), 450–468 (2012)
2. Bertero, M., Boccacci, P.: *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, Bristol (1998)
3. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. *Acta Numer.* **25**, 161–319 (2016)
4. di Serafino, D., Landi, G., Viola, M.: Directional TGV-based image restoration under Poisson noise. *J. Imaging* **7**(6), 99 (2021)
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
6. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2005)
7. Daubechies, I., Defrise, M., Mol, C.D.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **57**(11), 1413–1457 (2004)
8. Bonettini, S., Porta, F., Ruggiero, V.: A variable metric forward-backward method with extrapolation. *SIAM J. Sci. Comput.* **38**, 2558–2584 (2016)
9. Chouzenoux, E., Pesquet, J.-C., Repetti, A.: Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *J. Optim. Theory Appl.* **162**, 107–132 (2014)
10. Frankel, P., Garrigos, G., Peypouquet, J.: Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates. *J. Optim. Theory Appl.* **165**(3), 874–900 (2015)
11. Ghanbari, H., Scheinberg, K.: Proximal quasi-Newton methods for regularized convex optimization with linear and accelerated sublinear convergence rates. *Comput. Optim. Appl.* **69**, 597–627 (2018)

12. Lee, C., Wright, S.J.: Inexact successive quadratic approximation for regularized optimization. *Comput. Optim. Appl.* **72**, 641–674 (2019)
13. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **18**(11), 2419–34 (2009)
14. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for non-convex optimization. *SIAM J. Imaging Sci.* **7**(2), 1388–1419 (2014)
15. Polson, N.G., Scott, J.G., Willard, B.T.: Proximal algorithms in statistics and machine learning. *Stat. Sci.* **30**(4), 559–581 (2015)
16. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
17. Malitsky, Y., Pock, T.: A first-order primal-dual algorithm with linesearch. *SIAM J. Optim.* **28**(1), 411–432 (2018)
18. Chambolle, A., Delplancke, C., Ehrhardt, M.J., Schönlieb, C.-B., Tang, J.: Stochastic primal-dual hybrid gradient algorithm with adaptive step sizes. *J. Math. Imaging Vis.* **66**, 294–313 (2024)
19. Schmidt, M., Roux, N.L., Bach, F.: Convergence rates of inexact proximal-gradient methods for convex optimization. (2011)
20. Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and inexact forward-backward algorithms. *SIAM J. Optim.* **23**(3), 1607–1633 (2013)
21. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2011)
22. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* **2**, 17–40 (1976)
23. Wang, Y., Yin, W., Zeng, J.: Global convergence of ADMM in nonconvex nonsmooth optimization. *J. Sci. Comput.* **78**, 29–63 (2019)
24. Lee, J.D., Sun, Y., Saunders, M.A.: Proximal Newton-type methods for minimizing composite functions. *SIAM J. Optim.* **24**(3), 1420–1443 (2014)
25. Yue, M.-C., Zhou, Z., So, A.M.C.: A family of inexact SQA methods for non-smooth convex minimization with provable convergence guarantees based on the Luo-Tseng error bound property. *Math. Program.* **174**, 327–358 (2019)
26. Jiang, K., Sun, D., Toh, K.-C.: An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP. *SIAM J. Optim.* **22**(3), 1042–1064 (2012)
27. Kanzow, C., Lechner, T.: Efficient regularized proximal quasi-Newton methods for large-scale non-convex composite optimization problems. (2022)
28. Bonettini, S., Porta, F., Ruggiero, V., Zanni, L.: Variable metric techniques for forward-backward methods in imaging. *J. Comput. Appl. Math.* **385**, 113192 (2021)
29. Bonettini, S., Rebegoldi, S., Ruggiero, V.: Inertial variable metric techniques for the inexact forward-backward algorithm. *SIAM J. Sci. Comput.* **40**(5), 3180–3210 (2018)
30. Lantéri, H., Roche, M., Cuevas, O., Aime, C.: A general method to devise maximum likelihood signal restoration multiplicative algorithms with non-negativity constraints. *Signal Process.* **81**(5), 945–974 (2001)
31. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math. Dokl.* **269**, 543–547 (1983)
32. Chen, J., Loris, I.: On starting and stopping criteria for nested primal-dual iterations. *Numer. Algorithms* **82**, 605–621 (2019)
33. Rebegoldi, S., Calatroni, L.: Scaled, inexact and adaptive generalized FISTA for strongly convex optimization. *SIAM J. Optim.* **32**(3), 2428–2459 (2022)
34. Bonettini, S., Prato, M., Rebegoldi, S.: A nested primal-dual FISTA-like scheme for composite convex optimization problems. *Comput. Optim. Appl.* **84**, 85–123 (2023)
35. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *J. Phys. D.* **60**(1–4), 259–268 (1992)
36. Huang, J., Donatelli, M., Chan, R.: Nonstationary iterated thresholding algorithms for image deblurring. *Inverse Probl. Imaging* **7**(3), 717–736 (2013)
37. Donatelli, M., Hanke, M.: Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring. *Inverse Probl.* **29**(9), 095008 (2013)
38. Cai, Y., Donatelli, M., Bianchi, D., Huang, T.-Z.: Regularization preconditioners for frame-based image deblurring with reduced boundary artifacts. *SIAM J. Sci. Comput.* **38**(1), 164–189 (2016)
39. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)

40. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **93**, 273–299 (1965)
41. Polyak, B.: *Introduction to Optimization*. Optimization Software - Inc., Newyork (1987)
42. Combettes, P.L., Vu, B.C.: Variable metric quasi-Fejér monotonicity. *Nonlinear Anal.* **78**, 17–31 (2013)
43. Ciarlet, P.G.: *Introduction to Numerical Linear Algebra and Optimisation*. Cambridge University Press, Cambridge (1989)
44. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: *Proceedings of the 2011 International Conference on Computer Vision*, vol. 1, pp 1762–1769 (2011)
45. Chambolle, A., Dossal, C.: On the convergence of the iterates of the “fast iterative shrinkage/thresholding algorithm”. *J. Optim. Theory Appl.* **166**(3), 968–982 (2015)
46. Hansen, P.C., Nagy, J.G., O’Leary, D.P.: *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, Philadelphia (2006)
47. Liu, J., Huang, T.-Z., Selesnick, I.W., Lv, X.-G., Chen, P.-Y.: Image restoration using total variation with overlapping group sparsity. *Inform. Sci.* **295**, 232–246 (2015)
48. Donatelli, M.: On nondecreasing sequences of regularization parameters for nonstationary iterated Tikhonov. *Numer. Algorithms* **60**, 651–668 (2012)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.