On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems *

Radu Ioan Boţ †   Ernö Robert Csetnek ‡

August 25, 2014

Abstract. In this paper we analyze the convergence rate of the sequence of objective function values of a primal-dual proximal-point algorithm recently introduced in the literature for solving a primal convex optimization problem having as objective the sum of linearly composed infimal convolutions, nonsmooth and smooth convex functions and its Fenchel-type dual one. The theoretical part is illustrated by numerical experiments in image processing.

Key Words. convex optimization, minimization algorithm, duality, gap function, convergence rate, subdifferential

AMS subject classification. 47H05, 65K05, 90C25

1 Introduction and preliminaries

Due to their applications to fields like signal and image processing, support vector machines classification, multifacility location problems, clustering, network communication, portfolio optimization, etc., the numerical investigation in Hilbert spaces of nonsmooth optimization problems with intricate objective functions through proximal-point algorithms continues to attract the interest of many researchers (see [1, 2, 6–9, 11–14, 20]).

The proximal point algorithm [17] for minimizing a convex objective function is an excellent starting point for anyone interested in this topic. Since than, the number of works dedicated to this issue grown rapidly, due to handling of optimization problems having objective functions with more involved structure. Let us mention here the meanwhile classical *forward-backward algorithm* (see [1, 12]) and *forward-backward-forward (Tseng) algorithm* (see [1, 9, 19]), designed for solving convex optimization problems having as objective the sum of a proper, convex and lower semicontinuous (and not necessarily smooth) function with a convex differentiable one with Lipschitzian gradient. Here, a

*Dedicated to the founder and former editor-in-chief of the journal Optimization, Professor Karl-Heinz Elster, and to Professor Alfred Göpfert in celebration of his 80th birthday.

†University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at. Research partially supported by DFG (German Research Foundation), project BO 2516/4-1.

‡University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: ernoe.robert.csetnek@univie.ac.at. Research supported by DFG (German Research Foundation), project BO 2516/4-1.

backward step means that the nonsmooth function is evaluated in the iterative scheme via its proximal point, while a forward step means that the smooth function is evaluated via its gradient.

Lately, a new class of so-called *primal-dual splitting algorithms* has been intensively studied in the literature (see [6–9, 11, 13, 14, 20]). The iterative schemes in this family have the remarkable property that they concomitantly solve a primal convex optimization problem and its Fenchel-type dual and are able to handle optimization problems with intricate objectives, like the sum of linearly composed infimal convolutions of convex functions, nonsmooth and smooth convex functions. These methods have as further highlight that all the functions and operators that are present in the objective are evaluated individually in the algorithm.

For the primal-dual splitting algorithms mainly convergence statements for the sequence of iterates are available. However, especially from the point of view of solving real-life problems, the investigation of the convergence of the sequence of objective function is of equal importance (see [8, 11]).

It is the aim of this paper to investigate the convergence property of the sequence of objective function values of a primal-dual splitting algorithm recently introduced in [20] for convex optimization problems with intricate objective functions. By making use of the so-called primal-dual gap function attached to the structure of the problem, we are able to prove a convergence rate of order $\mathcal{O}(1/n)$. The results are formulated in the spirit of the ones given in [11] in a more particular setting.

The structure of the manuscript is as follows. The reminder of this section is dedicated to the introduction of some notations which are used throughout the paper and to the formulation of the primal-dual pair of convex optimization problems under investigation. In the next section we give the main result and discuss some of its consequences, while in Section 3 we illustrate the theoretical part by numerical experiments in image processing.

For the notations used we refer the reader to [1, 3, 4, 15, 18, 21]. Let $\mathcal{H}$ be a real Hilbert space with *inner product* $\langle\cdot,\cdot\rangle$ and associated *norm* $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. The symbols $\rightharpoonup$ and $\rightarrow$ denote weak and strong convergence, respectively. When $\mathcal{G}$ is another Hilbert space and $K : \mathcal{H} \rightarrow \mathcal{G}$ a linear continuous operator, then the *norm* of K is defined as $\|K\| = \sup\{\|Kx\| : x \in \mathcal{H}, \|x\| \leq 1\}$, while $K^* : \mathcal{G} \rightarrow \mathcal{H}$, defined by $\langle K^*y, x\rangle = \langle y, Kx\rangle$ for all $(x, y) \in \mathcal{H} \times \mathcal{G}$, denotes the *adjoint operator* of K.

For a function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ is the extended real line, we denote by $\operatorname{dom} f = \{x \in \mathcal{H} : f(x) < +\infty\}$ its *effective domain* and say that f is *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. We denote by $\Gamma(\mathcal{H})$ the family of proper convex and lower semi-continuous extended real-valued functions defined on $\mathcal{H}$. Let $f^* : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, $f^*(u) = \sup_{x\in\mathcal{H}}\{\langle u, x\rangle - f(x)\}$ for all $u \in \mathcal{H}$, be the *conjugate function* of f. The *subdifferential* of f at $x \in \mathcal{H}$, with $f(x) \in \mathbb{R}$, is the set $\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x\rangle \ \forall y \in \mathcal{H}\}$. We take by convention $\partial f(x) := \emptyset$, if $f(x) \in \{\pm\infty\}$. We denote by $\operatorname{ran}(\partial f) = \cup_{x\in\mathcal{H}}\partial f(x)$ the *range* of the subdifferential operator. The operator $(\partial f)^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $x \in (\partial f)^{-1}u$ if and only if $u \in \partial f(x)$. Notice that in case $f \in \Gamma(\mathcal{H})$ we have $(\partial f)^{-1} = \partial f^*$. For $f, g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ two proper functions, we consider their *infimal convolution*, which is the function $f \Box g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, defined by $(f \Box g)(x) = \inf_{y\in\mathcal{H}}\{f(y) + g(x - y)\}$, for all $x \in \mathcal{H}$. Further, the *parallel sum* of the subdifferential operators $\partial f, \partial g : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $\partial f \Box \partial g : \mathcal{H} \rightrightarrows \mathcal{H}, \partial f \Box \partial g = ((\partial f)^{-1} + (\partial g)^{-1})^{-1}$. In case $f, g \in \Gamma(\mathcal{H})$ and a regularity condition is fulfilled, according to [1, Proposition 24.27] we have $\partial f \Box \partial g = \partial(f \Box g)$, and this justifies the notation used for the parallel sum.

Let $S \subseteq \mathcal{H}$ be a nonempty set. The *indicator function* of S, $\delta_S : \mathcal{H} \to \overline{\mathbb{R}}$, is the function which takes the value 0 on S and $+\infty$ otherwise. The subdifferential of the indicator function is the *normal cone* of S, that is $N_S(x) = \{u \in \mathcal{H} : \langle u, y - x\rangle \leq 0 \ \forall y \in S\}$, if $x \in S$ and $N_S(x) = \emptyset$ for $x \notin S$.

When $f \in \Gamma(\mathcal{H})$ and $\gamma > 0$, for every $x \in \mathcal{H}$ we denote by $\operatorname{prox}_{\gamma f}(x)$ the *proximal point* of parameter γ of f at x, which is the unique optimal solution of the optimization problem

$$\inf_{y\in\mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{1}$$

Let us mention that $\operatorname{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$, called *proximal mapping*, is a single-valued operator fulfilling the extended *Moreau's decomposition formula*

$$\operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{(1/\gamma)f^*} \circ \gamma^{-1} \operatorname{Id}_{\mathcal{H}} = \operatorname{Id}_{\mathcal{H}}. \tag{2}$$

We notice that for $f = \delta_S$, where $S \subseteq \mathcal{H}$ is a nonempty convex and closed set, it holds

$$\operatorname{prox}_{\gamma\delta_S} = P_S, \tag{3}$$

where $P_S : \mathcal{H} \to C$ denotes the *orthogonal projection operator* on S (see [1, Example 23.3 and Example 23.4]). Finally, let us recall that the function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is said to be γ*-strongly convex* for $\gamma > 0$, if $f - \frac{\gamma}{2}\|\cdot\|^2$ is a convex function.

1.1 Problem formulation

The starting point of our investigation is the following problem.

Problem 1 *Let $\mathcal{H}$ be a real Hilbert space, $z \in \mathcal{H}$, $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ a convex and differentiable function with a η^{-1}-Lipschitz continuous gradient for $\eta > 0$. Let m be a strictly positive integer and for $i = 1, ..., m$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i \in \mathcal{G}_i$, $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that l_i is ν_i-strongly convex for $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ a nonzero linear continuous operator. Consider the convex optimization problem*

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \Box l_i)(L_i x - r_i) + h(x) - \langle x, z\rangle \right\} \tag{4}$$

and its Fenchel-type dual problem

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m} \left\{ -(f^* \Box h^*)\left(z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i\rangle \right) \right\}. \tag{5}$$

By employing the classical forward-backward algorithm (see [12]) in a renormed product space, Vũ proposed in [20] an iterative scheme for solving a slightly modified version of Problem 1 formulated in the presence of weights $w_i \in (0, 1]$, $i = 1,..., m$, with $\sum_{i=1}^{m} w_i = 1$ for the terms occurring in the second summand of the primal optimization problem and with the corresponding dual. The following result is an adaption of [20, Theorem 3.1] to Problem 1 to the error-free case and when $\lambda_n = 1$ for any $n \geq 0$.

Theorem 2 *In Problem 1 suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^*\big((\partial g_i \square \partial l_i)(L_i \cdot - r_i)\big) + \nabla h\right). \tag{6}$$

Let τ and σ_i, $i = 1,..., m$, be strictly positive numbers such that

$$2 \cdot \min\{\tau^{-1}, \sigma_1^{-1},..., \sigma_m^{-1}\} \cdot \min\{\eta, \nu_1,..., \nu_m\} \cdot \left(1 - \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}\right) > 1. \tag{7}$$

Let $(x_0, v_{1,0},..., v_{m,0}) \in \mathcal{H}\times \mathcal{G}_1 \times ...\times \mathcal{G}_m$ and for any $n \geq 0$ set:

$$\begin{aligned}
&x_{n+1} = \operatorname{prox}_{\tau f}\left[x_n - \tau\left(\textstyle\sum_{i=1}^{m} L_i^* v_{i,n} + \nabla h(x_n) - z\right)\right]\\
&y_n = 2x_{n+1} - x_n\\
&v_{i,n+1} = \operatorname{prox}_{\sigma_i g_i^*}[v_{i,n} + \sigma_i(L_i y_n - \nabla l_i^*(v_{i,n}) - r_i)],\ i = 1,..., m.
\end{aligned}$$

Then the following statements are true:

(a) there exist $\overline{x} \in \mathcal{H}$, an optimal solution to (4), and $(\overline{v}_1,..., \overline{v}_m) \in \mathcal{G}_1 \times ...\times \mathcal{G}_m$, an optimal solution to (5), such that the optimal objective values of the two problems coincide, the optimality conditions

$$z - \sum_{i=1}^{m} L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{v}_i \in (\partial g_i \square \partial l_i)(L_i \overline{x} - r_i),\ i = 1,..., m \tag{8}$$

are fulfilled and $x_n \rightharpoonup \overline{x}$ and $(v_{1,n},..., v_{m,n}) \rightharpoonup (\overline{v}_1,..., \overline{v}_m)$ as $n \to +\infty$.

(b) if h is strongly convex, then $x_n \to \overline{x}$ as $n \to +\infty$.

(c) if l_i^* is strongly convex for some $i \in \{1, ..., m\}$, then $v_{i,n} \to \overline{v}_i$ as $n \to +\infty$.

Notice that the research in [20] is closely related to [11] and [14], where the solving of primal-dual pairs of convex optimization problems by proximal splitting methods is considered, as well. More exactly, the convergence property of [11, Algorithm 1] proved in [11, Theorem 1] follow as special instance of the main result in [20]. On the other hand, Condat proposes [14] an algorithm which is also an extension of the one given in [11]. Let us also mention that two variants of the algorithm in Theorem 2 (one of them under the use of variable step sizes) have been proposed in [7] for which it was possible to determine convergence rates for the sequence of iterates.

Before we proceed, some comments are in order.

Remark 3 The function l_i^* is Fréchet differentiable with ν_i^{-1}-Lipschitz continuous gradient for $i = 1,..., m$ (see [1, Theorem 18.15], [21, Corollary 3.5.11, Remark 3.5.3]), hence the use of the gradient of l_i^* in the algorithm makes sense.

Remark 4 If $(\overline{x}, \overline{v}_1,..., \overline{v}_m) \in \mathcal{H}\times \mathcal{G}_1 \times ... \times \mathcal{G}_m$ satisfies the conditions in (8), then $\overline{x}$ is an optimal solution of (4), $(\overline{v}_1,..., \overline{v}_m)$ is an optimal solution of (5) and the optimal objective values of the two problems coincide.

In case a regularity condition is fulfilled, the optimality conditions (8) are also necessary. More precisely, if the primal problem (4) has an optimal solution $\overline{x}$ and a suitable regularity condition is fulfilled, then relation (6) holds and there exists $(\overline{v}_1,..., \overline{v}_m)$, an optimal solution to (5), such that $(\overline{x}, \overline{v}_1,..., \overline{v}_m)$ satisfies the optimality conditions (8).

For the readers convenience we present some regularity conditions which are suitable in this context. One of the weakest qualification conditions of interiority-type reads (see, for instance, [13, Proposition 4.3, Remark 4.4], [21, Theorem 2.8.3 (vii)])

$$(r_1,\dots,r_m) \in \operatorname{sqri}\left(\prod_{i=1}^{m}(\operatorname{dom} g_i + \operatorname{dom} l_i) - \{(L_1x,\dots,L_mx) : x \in \operatorname{dom} f\}\right). \tag{9}$$

Here, for $\mathcal{H}$ a real Hilbert space and $S \subseteq \mathcal{H}$ a convex set, we denote by

$$\operatorname{sqri} S := \{x \in S : \cup_{\lambda>0}\lambda(S-x) \text{ is a closed linear subspace of } \mathcal{H}\}$$

its *strong quasi-relative interior*. Notice that we always have $\operatorname{int} S \subseteq \operatorname{sqri} S$ (in general this inclusion may be strict). If $\mathcal{H}$ is finite-dimensional, then $\operatorname{sqri} S$ coincides with $\operatorname{ri} S$, the *relative interior* of S, which is the interior of S with respect to its affine hull. The condition (9) is fulfilled if (i) for any $i = 1,\dots,m$, $\operatorname{dom} g_i = \mathcal{G}_i$ or $\operatorname{dom} h_i = \mathcal{G}_i$ or (ii) $\mathcal{H}$ and $\mathcal{G}_i$ are finite-dimensional and there exists $x \in \operatorname{ri}\operatorname{dom} f$ such that $L_ix - r_i \in \operatorname{ri}\operatorname{dom} g_i + \operatorname{ri}\operatorname{dom} l_i$, $i = 1,\dots,m$ (see [13, Proposition 4.3]). For other regularity conditions we refer the reader to [1,3–5,21].

Remark 5 In the previous remark the existence of optimal solutions for problem (4) has been considered as a hypothesis for being able to formulate optimality conditions. Next we will discuss some conditions ensuring the existence of a primal optimal solution. Suppose that the primal problem (4) is feasible, which means that its optimal objective value is not identical $+\infty$. The existence of optimal solutions for (4) is guaranteed if for instance, $f + h + \langle\cdot, -z\rangle$ is coercive (that is $\lim_{\|x\|\to\infty}(f + h + \langle\cdot, -z\rangle)(x) = +\infty$) and for any $i = 1,\dots,m$, g_i is bounded from below. Indeed, under these circumstances, the objective function of (4) is coercive (use also [1, Corollary 11.16 and Proposition 12.14] to show that for any $i = 1,\dots,m$, $g_i \Box l_i$ is bounded from below and $g_i \Box l_i \in \Gamma(\mathcal{G}_i)$) and the statement follows via [1, Corollary 11.15]. On the other hand, if $f + h$ is strongly convex, then the objective function of (4) is strongly convex, too, thus (4) has a unique optimal solution (see [1, Corollary 11.16]).

Remark 6 In case $z = 0$, $h \equiv 0$, $r_i = 0$ and $l_i = \delta_{\{0\}}$ for any $i = 1,\dots,m$, the optimization problems (4) and (5) become

$$\inf_{x\in\mathcal{H}}\left\{f(x) + \sum_{i=1}^{m}(g_i \circ L_i)(x)\right\} \tag{10}$$

and, respectively

$$\sup_{v_i\in\mathcal{G}_i,\, i=1,..,m}\left\{-f^*\left(-\sum_{i=1}^{m}L_i^*v_i\right) - \sum_{i=1}^{m}g_i^*(v_i)\right\}. \tag{11}$$

It is mentioned in [20, Remark 3.3] that the convergence results in Theorem 2 hold if one replaces (7) by the condition

$$\tau\sum_{i=1}^{m}\sigma_i\|L_i\|^2 < 1. \tag{12}$$

The convergence (of an equivalent form) of the algorithm obtained in this setting has been investigated also in [6]. Moreover, the case $m = 1$ has been addressed in [11].

2 Convergence rate

In the setting of Problem 1 we introduce for $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ given nonempty sets the *primal-dual gap function* $\mathcal{G}_{B_1,B_2} \colon \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m \to \overline{\mathbb{R}}$ defined by

$$
\begin{aligned}
&\mathcal{G}_{B_1,B_2}(x, v_1, \ldots, v_m) = \\
&\sup_{(v_1',\ldots,v_m') \in B_2} \left\{ \sum_{i=1}^m \langle L_i x - r_i, v_i' \rangle + f(x) + h(x) - \langle x, z \rangle - \sum_{i=1}^m \Big(g_i^*(v_i') + l_i^*(v_i') \Big) \right\} \\
&- \inf_{x' \in B_1} \left\{ \sum_{i=1}^m \langle L_i x' - r_i, v_i \rangle + f(x') + h(x') - \langle x', z \rangle - \sum_{i=1}^m \Big(g_i^*(v_i) + h_i^*(v_i) \Big) \right\} \\
&= f(x) + h(x) - \langle x, z \rangle + \sup_{(v_1',\ldots,v_m') \in B_2} \left[\sum_{i=1}^m \langle L_i x - r_i, v_i' \rangle - \sum_{i=1}^m \Big(g_i^*(v_i') + l_i^*(v_i') \Big) \right] \\
&- \left\{ - \sum_{i=1}^m \left(g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle \right) + \inf_{x' \in B_1} \left[\sum_{i=1}^m \langle L_i x', v_i \rangle + f(x') + h(x') - \langle x', z \rangle \right] \right\}.
\end{aligned}
$$

Remark 7 If we consider the primal-dual pair of convex optimization problems from Remark 6 in case $m = 1$, then the primal-dual gap function defined above is nothing else than the one introduced in [11].

Remark 8 The primal-dual gap function defined above has been used in [8] in order investigate the convergence rate for the sequence of objective function values for the primal-dual splitting algorithm of forward-backward-forward type proposed in [13]. If $(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ satisfies the optimality conditions (8), then $\mathcal{G}_{B_1,B_2}(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \geq 0$ (see also [8, 11]).

We are now able to state the main result of the paper.

Theorem 9 *In Problem 1 suppose that*

$$
z \in \operatorname{ran} \left(\partial f + \sum_{i=1}^m L_i^* \left((\partial g_i \Box \partial l_i)(L_i \cdot - r_i) \right) + \nabla h \right). \tag{13}
$$

Let τ and σ_i, $i = 1, \ldots, m$, be strictly positive numbers such that

$$
\min\{\tau^{-1}, \sigma_1^{-1}, \ldots, \sigma_m^{-1}\} \cdot \min\{\eta, \nu_1, \ldots, \nu_m\} \cdot \left(1 - \sqrt{\tau \sum_{i=1}^m \sigma_i \|L_i\|^2} \right) > 1. \tag{14}
$$

Let $(x_0, v_{1,0}, \ldots, v_{m,0}) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$ and for any $n \geq 0$ set:

$$
\begin{aligned}
&x_{n+1} = \operatorname{prox}_{\tau f} \left[x_n - \tau \left(\textstyle\sum_{i=1}^m L_i^* v_{i,n} + \nabla h(x_n) - z \right) \right] \\
&y_n = 2x_{n+1} - x_n \\
&v_{i,n+1} = \operatorname{prox}_{\sigma_i g_i^*} [v_{i,n} + \sigma_i (L_i y_n - \nabla l_i^*(v_{i,n}) - r_i)], \ i = 1, \ldots, m.
\end{aligned}
$$

Then the following statements are true:

(a) there exist $\overline{x} \in \mathcal{H}$, an optimal solution to (4), and $(\overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$, an optimal solution to (5), such that the optimal objective values of the two problems coincide, the optimality conditions

$$
z - \sum_{i=1}^m L_i^* \overline{v}_i \in \partial f(\overline{x}) + \nabla h(\overline{x}) \ \text{and} \ \overline{v}_i \in (\partial g_i \Box \partial l_i)(L_i \overline{x} - r_i), \ i = 1, \ldots, m \tag{15}
$$

are fulfilled and $x_n \rightharpoonup \overline{x}$ and $(v_{1,n},..., v_{m,n}) \rightharpoonup (\overline{v}_1,..., \overline{v}_m)$ as $n \to +\infty$;

(b) if $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \mathcal{G}_1 \times ... \times \mathcal{G}_m$ are nonempty bounded sets, then for $x^N = \frac{1}{N}\sum_{n=1}^{N} x_{n+1}$ and $v_i^N = \frac{1}{N}\sum_{n=1}^{N} v_{i,n}$ we have $(x^N, v_1^N, ..., v_m^N) \rightharpoonup (\overline{x}, \overline{v}_1, ..., \overline{v}_m)$ as $N \to +\infty$ and for any $N \geq 1$

$$\mathcal{G}_{B_1,B_2}(x^N, v_1^N, ..., v_m^N) \leq \frac{C(B_1, B_2)}{N},$$

where

$$\begin{aligned} C(B_1, B_2) \;=\; & \sup_{x \in B_1}\left\{\frac{1}{2\tau}\|x_1 - x\|^2\right\} + \frac{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{2\tau}\|x_1 - x_0\|^2 + \\ & \sup_{(v_1,...,v_m) \in B_2}\left\{\sum_{i=1}^{m} \frac{1}{2\sigma_i}\|v_{i,0} - v_i\|^2 + \sum_{i=1}^{m}\langle L_i(x_1 - x_0), v_{i,0} - v_i\rangle\right\}; \end{aligned}$$

(c) if g_i is Lipschitz continuous on $\mathcal{G}_i$ for any $i = 1, ..., m$, then for any $N \geq 1$ we have

$$\begin{aligned} 0 \;\leq\; & \left(f(x^N) + \sum_{i=1}^{m}(g_i \Box l_i)(L_i x^N - r_i) + h(x^N) - \langle x^N, z\rangle\right) \\ & - \left(f(\overline{x}) + \sum_{i=1}^{m}(g_i \Box l_i)(L_i \overline{x} - r_i) + h(\overline{x}) - \langle \overline{x}, z\rangle\right) \leq \frac{C(B_1, B_2)}{N} \end{aligned} \tag{16}$$

where B_1 is any bounded and weak sequentially closed set containing the sequence $(x_n)_{n\in\mathbb{N}}$ (which is the case if for instance B_1 is bounded, convex and closed with respect to the strong topology of $\mathcal{H}$ and contains the sequence $(x_n)_{n\in\mathbb{N}}$) and B_2 is any bounded set containing $\operatorname{dom} g_1^* \times ... \times \operatorname{dom} g_m^*$;

(d) if $\operatorname{dom} g_i + \operatorname{dom} l_i = \mathcal{G}_i$ for any $i = 1, ..., m$, and one of the following conditions is fulfilled:

(d1) $\mathcal{H}$ is finite-dimensional;

(d2) $\mathcal{G}_i$ is finite-dimensional for any $i = 1, ..., m$;

(d3) h is strongly convex;

then the inequality (16) holds for any $N \geq 1$, where B_1 is taken as in (c) and B_2 is any bounded set containing $\Pi_{i=1}^{m} \cup_{N \geq 1} \partial(g_i \Box l_i)(L_i x^N - r_i)$.

Proof. In order to allow the reader to follow much easier the proof of the main result of the paper we will give it for $m = 1$. The general case can be shown in a similar manner. Further, we denote $\mathcal{G} := \mathcal{G}_1$, $r := r_1$, $g := g_1$, $l := l_1$, $\nu := \nu_1$, $L := L_1$, $\sigma := \sigma_1$, $v_n := v_{1,n}$ for any $n \geq 0$, $v^N := v_1^N$ for any $N \geq 1$ and $\bar{v} := \bar{v}_1$. Hence, τ and σ are strictly positive numbers fulfilling that

$$\min\{\tau^{-1}, \sigma^{-1}\} \cdot \min\{\eta, \nu\} \cdot \left(1 - \sqrt{\tau\sigma\|L\|^2}\right) > 1, \tag{17}$$

while for $(x_0, v_0) \in \mathcal{H} \times \mathcal{G}$ and for any $n \geq 0$ the iterative scheme reads:

$$\begin{aligned} & x_{n+1} = \operatorname{prox}_{\tau f}\left[x_n - \tau\left(L^* v_n + \nabla h(x_n) - z\right)\right] \\ & y_n = 2x_{n+1} - x_n \\ & v_{n+1} = \operatorname{prox}_{\sigma g^*}[v_n + \sigma(L y_n - \nabla l^*(v_n) - r)]. \end{aligned}$$

(a) The statement is a direct consequence of Theorem 2, since condition (14) implies (7).

(b) The fact that $(x^N, v^N) \rightharpoonup (\overline{x}, \overline{v})$ as $N \to +\infty$ follows from the Stolz-Cesàro Theorem. Let us show now the inequality concerning the primal-dual gap function. To this aim we fix for the beginning $n \geq 0$.

From the definition of the iterates we derive

$$\frac{1}{\tau}(x_{n+1} - x_{n+2}) - (L^* v_{n+1} + \nabla h(x_{n+1}) - z) \in \partial f(x_{n+2}),$$

hence the definition of the subdifferential delivers the inequality

$$\begin{aligned} f(x) \geq \quad & f(x_{n+2}) + \frac{1}{\tau}\langle x_{n+1} - x_{n+2}, x - x_{n+2}\rangle - \langle L^* v_{n+1} - z, x - x_{n+2}\rangle \\ & -\langle \nabla h(x_{n+1}), x - x_{n+2}\rangle \ \forall x \in \mathcal{H}. \end{aligned} \tag{18}$$

Similarly, we deduce

$$\frac{1}{\sigma}(v_n - v_{n+1}) + Ly_n - \nabla l^*(v_n) - r \in \partial g^*(v_{n+1}),$$

hence

$$\begin{aligned} g^*(v) \ \geq \quad & g^*(v_{n+1}) + \frac{1}{\sigma}\langle v_n - v_{n+1}, v - v_{n+1}\rangle + \langle Ly_n - r, v - v_{n+1}\rangle \\ & -\langle \nabla l^*(v_n), v - v_{n+1}\rangle \ \forall v \in \mathcal{G}. \end{aligned} \tag{19}$$

We claim that

$$h(x) - h(x_{n+2}) - \langle \nabla h(x_{n+1}), x - x_{n+2}\rangle \geq -\frac{\eta^{-1}}{2}\|x_{n+2} - x_{n+1}\|^2 \ \forall x \in \mathcal{H}. \tag{20}$$

Indeed, we have

$$\begin{aligned} & h(x) - h(x_{n+2}) - \langle \nabla h(x_{n+1}), x - x_{n+2}\rangle \\ \geq \ & h(x_{n+1}) + \langle \nabla h(x_{n+1}), x - x_{n+1}\rangle - h(x_{n+2}) - \langle \nabla h(x_{n+1}), x - x_{n+2}\rangle \\ = \ & h(x_{n+1}) - h(x_{n+2}) + \langle \nabla h(x_{n+1}), x_{n+2} - x_{n+1}\rangle \\ \geq \ & -\frac{\eta^{-1}}{2}\|x_{n+2} - x_{n+1}\|^2, \end{aligned}$$

where the first inequality holds since h is convex and the second one follows by choosing $x = x_{n+1}$ and $y = x_{n+2}$ in the inequality

$$h(y) \leq h(x) + \langle \nabla h(x), y - x\rangle + \frac{\eta^{-1}}{2}\|y - x\|^2,$$

which is valid for an arbitrary continuously differentiable function with η^{-1}-Lipschitz continuous gradient (see [16, Lemma 1.2.3]). Hence (20) holds.

Similarly one can prove that

$$l^*(v) - l^*(v_{n+1}) - \langle \nabla l^*(v_n), v - v_{n+1}\rangle \geq -\frac{\nu^{-1}}{2}\|v_{n+1} - v_n\|^2 \ \forall v \in \mathcal{G}. \tag{21}$$

By adding the inequalities (18), (19), (20), (21) and noticing that

$$\langle x_{n+1} - x_{n+2}, x - x_{n+2}\rangle = -\frac{\|x_{n+1} - x\|^2}{2} + \frac{\|x_{n+1} - x_{n+2}\|^2}{2} + \frac{\|x_{n+2} - x\|^2}{2}$$

and

$$\langle v_n - v_{n+1}, v - v_{n+1}\rangle = -\frac{\|v_n - v\|^2}{2} + \frac{\|v_{n+1} - v_n\|^2}{2} + \frac{\|v_{n+1} - v\|^2}{2},$$

we deduce that for all $(x, v) \in \mathcal{H} \times \mathcal{G}$

$$\begin{aligned}
&\frac{\|x_{n+1} - x\|^2}{2\tau} + \frac{\|v_n - v\|^2}{2\sigma} \geq \\
&\frac{\|x_{n+2} - x\|^2}{2\tau} + \frac{\|v_{n+1} - v\|^2}{2\sigma} + \frac{1 - \eta^{-1}\tau}{2\tau}\|x_{n+1} - x_{n+2}\|^2 + \frac{1 - \nu^{-1}\sigma}{2\sigma}\|v_{n+1} - v_n\|^2 \\
&+ \Big(\langle Lx_{n+2} - r, v\rangle + f(x_{n+2}) + h(x_{n+2}) - \langle x_{n+2}, z\rangle - \big(g^*(v) + l^*(v)\big)\Big) \\
&- \Big(\langle Lx - r, v_{n+1}\rangle + f(x) + h(x) - \langle x, z\rangle - \big(g^*(v_{n+1}) + l^*(v_{n+1})\big)\Big) \\
&+ \langle L(x_{n+2} - y_n), v_{n+1} - v\rangle.
\end{aligned}$$

Taking into account the definition of y_n, we get the following estimation for the last term:

$$\begin{aligned}
&\langle L(x_{n+2} - y_n), v_{n+1} - v\rangle = \\
&\langle L(x_{n+2} - x_{n+1}), v_{n+1} - v\rangle - \langle L(x_{n+1} - x_n), v_n - v\rangle + \langle L(x_{n+1} - x_n), v_n - v_{n+1}\rangle \\
\geq &\langle L(x_{n+2} - x_{n+1}), v_{n+1} - v\rangle - \langle L(x_{n+1} - x_n), v_n - v\rangle \\
&- \left(\frac{\sigma\|L\|^2}{2\sqrt{\tau\sigma\|L\|^2}}\|x_{n+1} - x_n\|^2 + \frac{\sqrt{\tau\sigma\|L\|^2}}{2\sigma}\|v_{n+1} - v_n\|^2\right),
\end{aligned}$$

hence we obtain the inequality

$$\begin{aligned}
&\frac{\|x_{n+1} - x\|^2}{2\tau} + \frac{\|v_n - v\|^2}{2\sigma} \geq \\
&\frac{\|x_{n+2} - x\|^2}{2\tau} + \frac{\|v_{n+1} - v\|^2}{2\sigma} + \frac{1 - \eta^{-1}\tau}{2\tau}\|x_{n+1} - x_{n+2}\|^2 - \frac{\sqrt{\tau\sigma\|L\|^2}}{2\tau}\|x_{n+1} - x_n\|^2 \\
&+ \frac{1 - \nu^{-1}\sigma - \sqrt{\tau\sigma\|L\|^2}}{2\sigma}\|v_{n+1} - v_n\|^2 \\
&+ \Big(\langle Lx_{n+2} - r, v\rangle + f(x_{n+2}) + h(x_{n+2}) - \langle x_{n+2}, z\rangle - \big(g^*(v) + l^*(v)\big)\Big) \\
&- \Big(\langle Lx - r, v_{n+1}\rangle + f(x) + h(x) - \langle x, z\rangle - \big(g^*(v_{n+1}) + l^*(v_{n+1})\big)\Big) \\
&+ (\langle L(x_{n+2} - x_{n+1}), v_{n+1} - v\rangle - \langle L(x_{n+1} - x_n), v_n - v\rangle).
\end{aligned}$$

Summing up the above inequality from $n = 0$ to $N - 1$, where $N \in \mathbb{N}, N \geq 1$, we get

$$\begin{aligned}
&\frac{\|x_1 - x\|^2}{2\tau} + \frac{\|v_0 - v\|^2}{2\sigma} \geq \frac{\|x_{N+1} - x\|^2}{2\tau} + \frac{\|v_N - v\|^2}{2\sigma} \\
&+ \sum_{n=1}^{N-1} \frac{1 - \eta^{-1}\tau}{2\tau}\|x_{n+1} - x_n\|^2 + \frac{1 - \eta^{-1}\tau}{2\tau}\|x_N - x_{N+1}\|^2 - \frac{\sqrt{\tau\sigma\|L\|^2}}{2\tau}\sum_{n=1}^{N-1}\|x_{n+1} - x_n\|^2
\end{aligned}$$

$$
\begin{aligned}
&-\frac{\sqrt{\tau\sigma\|L\|^2}}{2\tau}\|x_1-x_0\|^2+\sum_{n=0}^{N-1}\frac{1-\nu^{-1}\sigma-\sqrt{\tau\sigma\|L\|^2}}{2\sigma}\|v_{n+1}-v_n\|^2\\
&+\sum_{n=1}^{N}\Big(\langle Lx_{n+1}-r,v\rangle+f(x_{n+1})+h(x_{n+1})-\langle x_{n+1},z\rangle-\big(g^*(v)+l^*(v)\big)\Big)\\
&-\sum_{n=1}^{N}\Big(\langle Lx-r,v_n\rangle+f(x)+h(x)-\langle x,z\rangle-\big(g^*(v_n)+l^*(v_n)\big)\Big)\\
&+(\langle L(x_{N+1}-x_N),v_N-v\rangle-\langle L(x_1-x_0),v_0-v\rangle).
\end{aligned}
$$

Further, for the last term we use the estimate (notice that $1-\eta^{-1}\tau>0$ due to (17))

$$
\langle L(x_{N+1}-x_N),v_N-v\rangle\geq-\left(\frac{(1-\eta^{-1}\tau)\sigma\|L\|^2}{2\tau\sigma\|L\|^2}\|x_{N+1}-x_N\|^2+\frac{\tau\sigma\|L\|^2}{2\sigma(1-\eta^{-1}\tau)}\|v_N-v\|^2\right)
$$

and conclude that for all $(x,v)\in\mathcal{H}\times\mathcal{G}$

$$
\begin{aligned}
&\frac{\|x_{N+1}-x\|^2}{2\tau}+\frac{1-\eta^{-1}\tau-\tau\sigma\|L\|^2}{2\sigma}\|v_N-v\|^2\\
&+\sum_{n=1}^{N-1}\frac{1-\eta^{-1}\tau-\sqrt{\tau\sigma\|L\|^2}}{2\tau}\|x_{n+1}-x_n\|^2+\sum_{n=0}^{N-1}\frac{1-\nu^{-1}\sigma-\sqrt{\tau\sigma\|L\|^2}}{2\sigma}\|v_{n+1}-v_n\|^2\\
&+\sum_{n=1}^{N}\Big(\langle Lx_{n+1}-r,v\rangle+f(x_{n+1})+h(x_{n+1})-\langle x_{n+1},z\rangle-\big(g^*(v)+l^*(v)\big)\Big)\\
&-\sum_{n=1}^{N}\Big(\langle Lx-r,v_n\rangle+f(x)+h(x)-\langle x,z\rangle-\big(g^*(v_n)+l^*(v_n)\big)\Big)\\
\leq&\frac{\|x_1-x\|^2}{2\tau}+\frac{\|v_0-v\|^2}{2\sigma}+\frac{\sqrt{\tau\sigma\|L\|^2}}{2\tau}\|x_1-x_0\|^2+\langle L(x_1-x_0),v_0-v\rangle.
\end{aligned}
$$

We can discard the first four terms in the left-hand side of the above inequality, since due to (17), we have

$$
1-\eta^{-1}\tau-\sqrt{\tau\sigma\|L\|^2}>0 \tag{22}
$$

and

$$
1-\nu^{-1}\sigma-\sqrt{\tau\sigma\|L\|^2}>0. \tag{23}
$$

Thus we obtain for all $(x,v)\in\mathcal{H}\times\mathcal{G}$ that

$$
\begin{aligned}
&\sum_{n=1}^{N}\Big(\langle Lx_{n+1}-r,v\rangle+f(x_{n+1})+h(x_{n+1})-\langle x_{n+1},z\rangle-\big(g^*(v)+l^*(v)\big)\Big)\\
&-\sum_{n=1}^{N}\Big(\langle Lx-r,v_n\rangle+f(x)+h(x)-\langle x,z\rangle-\big(g^*(v_n)+l^*(v_n)\big)\Big)\\
\leq&\frac{\|x_1-x\|^2}{2\tau}+\frac{\|v_0-v\|^2}{2\sigma}+\frac{\sqrt{\tau\sigma\|L\|^2}}{2\tau}\|x_1-x_0\|^2+\langle L(x_1-x_0),v_0-v\rangle.
\end{aligned}
$$

The conclusion follows by dividing by N both terms of the previous inequality, taking into account the definition of (x^N,v^N) and the convexity of the functions f,h and g^*,l^*, then passing to supremum over $x\in B_1$ and $v\in B_2$.

(c) According to [3, Proposition 4.4.6], the set $\operatorname{dom} g^*$ is bounded. Since B_1 is weak sequentially closed and $x_n \rightharpoonup \overline{x}$ we have $\overline{x} \in B_1$. Let be $N \geq 1$ fixed. We get from (b) that

$$\begin{aligned}
&\frac{C(B_1,B_2)}{N} \geq \mathcal{G}_{B_1,B_2}(x^N,v^N)\\
&\geq f(x^N)+h(x^N)-\langle x^N,z\rangle+\sup_{v'\in\operatorname{dom} g^*}\Big\{\langle Lx^N-r,v'\rangle-(g^*(v')+l^*(v')\Big\}\\
&\quad-\Big(\langle L\overline{x}-r,v^N\rangle+f(\overline{x})+h(\overline{x})-\langle\overline{x},z\rangle-\big(g^*(v^N)+l^*(v^N)\big)\Big).
\end{aligned}$$

Further, it follows

$$\begin{aligned}
&\sup_{v'\in\operatorname{dom} g^*}\Big\{\langle Lx^N-r,v'\rangle-(g^*(v')+l^*(v')\Big\}=\sup_{v'\in\mathcal{G}}\Big\{\langle Lx^N-r,v'\rangle-(g^*(v')+l^*(v')\Big\}\\
&=(g^*+l^*)^*(Lx^N-r)=(g^{**}\Box l^{**})(Lx^N-r)=(g\Box l)(Lx^N-r),
\end{aligned}$$

where we used [1, Proposition 15.2] (notice that $\operatorname{dom} l^*=\mathcal{G}$) and the Fenchel-Moreau Theorem (see for example [1, Theorem 13.32]). Furthermore, the Young-Fenchel inequality (see [1, Proposition 13.13]) guarantees that

$$g^*(v^N)+l^*(v^N)-\langle L\overline{x}-r,v^N\rangle=(g\Box h)^*(v^N)-\langle L\overline{x}-r,v^N\rangle\geq-(g\Box l)(L\overline{x}-r)$$

and the conclusion follows.

(d) We notice first that each of the conditions (d1),(d2) and (d3) implies that

$$Lx^N\to L\overline{x} \text{ as } N\to+\infty. \tag{24}$$

Indeed, in case of (d1) we use that $x^N\rightharpoonup\overline{x}$ as $N\to+\infty$, in case (d2) that $Lx^N\rightharpoonup L\overline{x}$ as $N\to+\infty$ (which is a consequence of $x^N\rightharpoonup\overline{x}$ as $N\to+\infty$), while in the last case we appeal Theorem 2(b).

We show first that $\cup_{N\geq1}\partial(g\Box l)(Lx^N-r)$ is a nonempty bounded set. The function $g\Box l$ belongs to $\Gamma(\mathcal{H})$, as already mentioned above. Further, as $\operatorname{dom}(g\Box l)=\operatorname{dom} g+\operatorname{dom} l=\mathcal{G}$, it follows that $g\Box l$ is everywhere continuous (see [1, Corollary 8.30]) and, consequently, everywhere subdifferentiable (see [1, Proposition 16.14(iv)]). Hence the claim concerning the nonemptiness of the set $\cup_{N\geq1}\partial(g\Box l)(Lx^N-r)$ is true. Moreover, since the subdifferential of $g\Box l$ is locally bounded at $L\overline{x}-r$ (see [1, Proposition 16.14(iii)]) and $Lx^N-r\to L\overline{x}-r$ as $N\to+\infty$ we easily derive from [1, Proposition 16.14(iii) and (ii)] that the set $\cup_{N\geq1}\partial(g\Box l)(Lx^N-r)$ is bounded.

Now we prove that the inequality (16) holds. Similarly as in (c), we have

$$\begin{aligned}
&\frac{C(B_1,B_2)}{N}\geq\mathcal{G}_{B_1,B_2}(x^N,v^N)\\
&\geq f(x^N)+h(x^N)-\langle x^N,z\rangle+\sup_{v'\in\cup_{N'\geq1}\partial(g\Box l)(Lx^{N'}-r)}\Big\{\langle Lx^N-r,v'\rangle-(g^*(v')+l^*(v')\Big\}\\
&\quad-\Big(\langle L\overline{x}-r,v^N\rangle+f(\overline{x})+h(\overline{x})-\langle\overline{x},z\rangle-\big(g^*(v^N)+l^*(v^N)\big)\Big).
\end{aligned}$$

Further, for any $N\geq1$ we have

$$\begin{aligned}
&\sup_{v'\in\cup_{N'\geq1}\partial(g\Box l)(Lx^{N'}-r)}\Big\{\langle Lx^N-r,v'\rangle-(g^*(v')+l^*(v')\Big\}\\
&\geq\sup_{v'\in\partial(g\Box l)(Lx^N-r)}\Big\{\langle Lx^N-r,v'\rangle-(g^*(v')+l^*(v')\Big\}=(g\Box l)(Lx^N-r),
\end{aligned}$$

where the last equality follows since $\partial(g\Box l)(Lx^N - r) \neq \emptyset$ via

$$\langle Lx^N - r, v' \rangle - (g^*(v') + l^*(v')) = \langle Lx^N - r, v' \rangle - (g\Box h)^*(v') = (g\Box l)(Lx^N - r),$$

which holds for every $v' \in \partial(g\Box l)(Lx^N - r)$ (see [1, Proposition 16.9]).

Using the same arguments as at the end of the proof of statement (c), the conclusion follows. ■

Remark 10 The conclusion of the above theorem remains true if condition (14) is replaced by (7), (22) and (23). Moreover, if one works in the setting of Remark 6, one can show that the conclusion of Theorem 9 remains valid if instead of (14) one assumes (12).

Remark 11 Let us mention that in Theorem 9(c) and (d) one can chose for B_1 any bounded set containing $\overline{x}$.

Remark 12 If f is Lipschitz continuous, then, similarly to Theorem 9(c), one can prove via Theorem 9(b) a convergence rate of order $\mathcal{O}(1/n)$ for the sequence of values of the objective function of the dual problem (5). The same conclusion follows in case f has full domain and one of the conditions (d1), (d2) and (d3') is fulfilled, where (d3') assumes that l_i^* is strongly convex for any $i = 1, ..., m$.

3 Numerical experiments in imaging

The aim of this section is to illustrate the theoretical results obtained in the previous section by means of two problems occurring in imaging. For the applications discussed in this section the images have been normalized, in order to make their pixels range in the closed interval from 0 to 1.

3.1 TV-L2-based image deblurring

The first numerical experiment that we consider concerns addresses an ill-conditioned linear inverse problem which arises in image deblurring. For a given matrix $A \in \mathbb{R}^{n \times n}$ describing a blur operator and a given vector $b \in \mathbb{R}^n$ representing the blurred and noisy image, the task is to estimate the unknown original image $\overline{x} \in \mathbb{R}^n$ fulfilling

$$A\overline{x} = b.$$

To this end we solve the following regularized convex minimization problem

$$\inf_{x \in [0,1]^n} \left\{ \|Ax - b\|_1 + \lambda (TV_{\text{iso}}(x) + \|x\|^2) \right\}, \tag{25}$$

where $\lambda > 0$ is a regularization parameter and $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$ is the discrete isotropic total variation functional. In this context, $x \in \mathbb{R}^n$ represents the vectorized image $X \in \mathbb{R}^{M \times N}$, where $n = M \cdot N$ and $x_{i,j}$ denotes the normalized value of the pixel located in the i-th row and the j-th column, for $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

The *isotropic total variation* $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$TV_{\text{iso}}(x) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} + \sum_{i=1}^{M-1} |x_{i+1,N} - x_{i,N}| + \sum_{j=1}^{N-1} |x_{M,j+1} - x_{M,j}|.$$

We show first that the optimization problem (25) can be written in the framework of Problem 1.

We denote $\mathcal{Y} = \mathbb{R}^n \times \mathbb{R}^n$ and define the linear operator $L : \mathbb{R}^n \to \mathcal{Y}$, $x_{i,j} \mapsto (L_1 x_{i,j}, L_2 x_{i,j})$, where

$$L_1 x_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j}, & \text{if } i < M \\ 0, & \text{if } i = M \end{cases} \text{ and } L_2 x_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases}.$$

The operator L represents a discretization of the gradient using reflexive (Neumann) boundary conditions and standard finite differences and fulfills $\|L\|^2 \leq 8$. For the formula for its adjoint operator $L^* : \mathcal{Y} \to \mathbb{R}^n$ we refer to [10].

For (y, z), $(p, q) \in \mathcal{Y}$, we introduce the inner product

$$\langle (y, z), (p, q) \rangle = \sum_{i=1}^{M} \sum_{j=1}^{N} y_{i,j} p_{i,j} + z_{i,j} q_{i,j}$$

and define $\|(y, z)\|_\times = \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{y_{i,j}^2 + z_{i,j}^2}$. One can check that $\|\cdot\|_\times$ is a norm on $\mathcal{Y}$ and that for every $x \in \mathbb{R}^n$ it holds $TV_{\text{iso}}(x) = \|Lx\|_\times$. The conjugate function $(\|\cdot\|_\times)^* : \mathcal{Y} \to \overline{\mathbb{R}}$ of $\|\cdot\|_\times$ is for every $(p, q) \in \mathcal{Y}$ given by

$$(\|\cdot\|_\times)^*(p, q) = \begin{cases} 0, & \text{if } \|(p, q)\|_{\times *} \leq 1 \\ +\infty, & \text{otherwise} \end{cases},$$

where

$$\|(p, q)\|_{\times *} = \sup_{\|(y,z)\|_\times \leq 1} \langle (p, q), (y, z) \rangle = \max_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \sqrt{p_{i,j}^2 + q_{i,j}^2}.$$

Therefore, the optimization problem (25) can be written in the form of

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g_1(Ax) + g_2(Lx) + h(x)\},$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f(x) = \delta_{[0,1]^n}(x)$, $g_1 : \mathbb{R}^n \to \mathbb{R}$, $g_1(y) = \|y - b\|_1$, $g_2 : \mathcal{Y} \to \mathbb{R}$, $g_2(y, z) = \lambda \|(y, z)\|_\times$ and $h : \mathbb{R}^n \to \mathbb{R}$, $h(x) = \lambda \|x\|^2$ (notice that the functions l_i are taken to be $\delta_{\{0\}}$ for $i = 1, 2$). For every $p \in \mathbb{R}^n$, it holds $g_1^*(p) = \delta_{[-1,1]^n}(p) + p^T b$, while for every $(p, q) \in \mathcal{Y}$, we have $g_2^*(p, q) = \delta_S(p, q)$, with $S = \{(p, q) \in \mathcal{Y} : \|(p, q)\|_{\times *} \leq \lambda\}$. Moreover, h is differentiable with $\eta^{-1} := 2\lambda$-Lipschitz continuous gradient. We solved this problem by the algorithm considered in this paper and to this end we made use of the following formulae

$$\text{prox}_{\gamma f}(x) = P_{[0,1]^n}(x) \ \ \forall x \in \mathbb{R}^n$$

$$\text{prox}_{\gamma g_1^*}(p) = P_{[-1,1]^n}(p - \gamma b) \ \ \forall p \in \mathbb{R}^n$$

(a) Original image (b) Blurred and noisy image (c) Reconstructed image

Figure 1: Figure (a) shows the original 256×256 boat test image, figure (b) shows the blurred and noisy image and figure (c) shows the averaged iterate generated by the algorithm after 400 iterations.

$$\operatorname{prox}_{\gamma g_2^*}(p,q) = P_S\,(p,q) \ \ \forall (p,q) \in \mathcal{Y},$$

where $\gamma > 0$ and the projection operator $P_S : \mathcal{Y} \to S$ is defined as (see [8])

$$(p_{i,j}, q_{i,j}) \mapsto \lambda \frac{(p_{i,j}, q_{i,j})}{\max\left\{\lambda, \sqrt{p_{i,j}^2 + q_{i,j}^2}\right\}}, \ 1 \leq i \leq M, \ 1 \leq j \leq N.$$

For the experiments we considered the 256×256 boat test image and constructed the blurred image by making use of a Gaussian blur operator of size 9×9 and standard deviation 4. In order to obtain the blurred and noisy image we added a zero-mean white Gaussian noise with standard deviation 10^{-3}. Figure 1 shows the original boat test image and the blurred and noisy one. It also shows the image reconstructed by the algorithm after 400 iterations in terms of the averaged iterate, when taking as regularization parameter $\lambda = 0.001$ and by choosing as parameters $\sigma_1 = 0.01, \sigma_2 = 0.7, \tau = 0.49$. On the other hand, in Figure 2 a comparison of the decrease of the objective function values is provided, in terms of the last and averaged iterates, underlying the rate of convergence of order $\mathcal{O}(1/n)$ for the latter.

3.2 TV-based image denoising

The second numerical experiment aims the solving of an image denoising problem via total variation regularization. More precisely, we deal with the convex optimization problem

$$\inf_{x \in \mathbb{R}^n} \left\{ \lambda\, TV_{\text{aniso}}(x) + \frac{1}{2}\|x - b\|^2 \right\}, \tag{26}$$

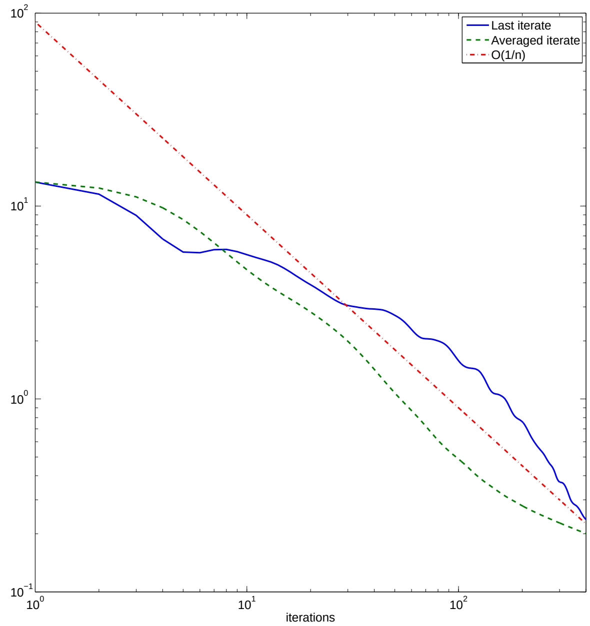

Figure 2: The figure shows the relative error in terms of function values for both the last and the averaged iterate generated by the algorithm after 400 iterations.

where $b \in \mathbb{R}^n$ is the observed noisy image, $\lambda > 0$ is the regularization parameter and $TV_{\text{aniso}} : \mathbb{R}^n \to \mathbb{R}$ is the discrete *anisotropic total variation* functional defined by

$$TV_{\text{aniso}}(x) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}| + \sum_{i=1}^{M-1} |x_{i+1,N} - x_{i,N}| + \sum_{j=1}^{N-1} |x_{M,j+1} - x_{M,j}|.$$

The optimization problem (26) can be equivalently written as

$$\inf_{x \in \mathbb{R}^n} \{g(Lx) + h(x)\},$$

where L and $\mathcal{Y}$ are the operator and the space, respectively, introduced in Section 3.1, $g : \mathcal{Y} \to \mathbb{R}$ is defined as $g(y_1, y_2) = \lambda \|(y_1, y_2)\|_1$ and $h : \mathbb{R}^n \to \mathbb{R}$, $h(x) = \frac{1}{2}\|x - b\|^2$ is differentiable with $\eta^{-1} := 1$-Lipschitz continuous gradient. Further, $g^* : \mathcal{Y} \to \overline{\mathbb{R}}$ is nothing else than

$$g^*(p_1, p_2) = (\lambda \| \cdot \|_1)^* (p_1, p_2) = \lambda \left\| \left(\frac{p_1}{\lambda}, \frac{p_2}{\lambda} \right) \right\|_1^* = \delta_{[-\lambda,\lambda]^n \times [-\lambda,\lambda]^n}(p_1, p_2),$$

hence

$$\text{prox}_{\gamma g^*}(p_1, p_2) = P_{[-\lambda,\lambda]^n \times [-\lambda,\lambda]^n}(p_1, p_2) \ \forall \gamma > 0 \ \forall (p_1, p_2) \in \mathcal{Y}.$$

For the experiments we used the 216×216 parrot test image. The noisy image was obtained after adding to the original image white Gaussian noise with standard deviation $\sigma = 0.12$.

(a) Original image (b) Noisy image (c) Reconstructed image

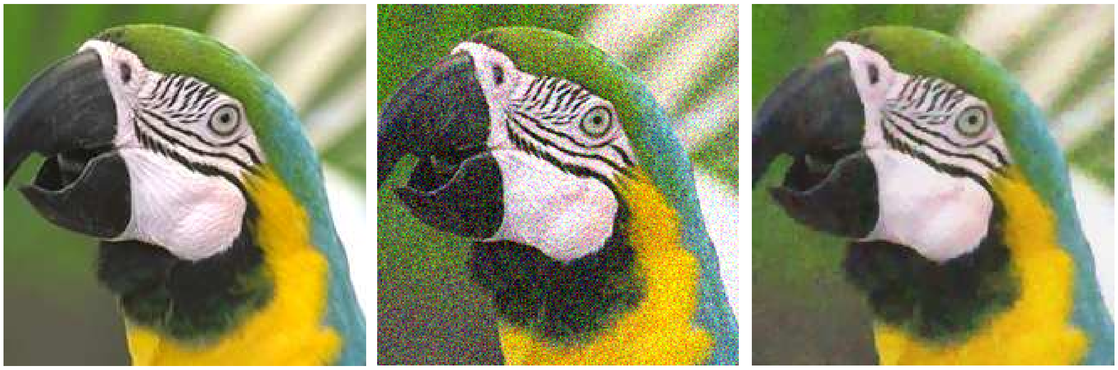

Figure 3: Figure (a) shows the original 216×216 parrot test image, figure (b) shows the noisy image and figure (c) shows the averaged iterate generated by the algorithm after 150 iterations.

Figure 3 shows the original, the noisy image and the image reconstructed by the algorithm after 150 iterations in terms of the averaged iterate and by using as parameters $\sigma_1 = 0.2, \tau = 0.479$ and $\lambda = 0.07$.

Let us notice that the function $\frac{1}{2}\|\cdot - b\|^2$ in the objective of the optimization problem (26) can be evaluated in the algorithm by a forward step via its gradient, but also by a backward step via its proximal mapping, which is given by

$$\operatorname{prox}_{\gamma h}(x) = \frac{1}{1+\gamma}(x + \gamma b)\ \forall \gamma > 0\ \forall x \in \mathbb{R}^n.$$

In Figure 4 we plotted the objective function values generated by the algorithm in Theorem 9 by evaluating h in both ways. We called the obtained plots (FB-f), when h was evaluated via its gradient (forward step), and (FB-b), when h was evaluated via its proximal mapping (backward step), respectively. We also solved the optimization problem (26), by dealing with h in a similar way, with the forward-backward-forward primal-dual algorithm introduced in [13], for the objective function values of which similar convergence properties have been reported (see [8]). In the case when h was evaluated via its gradient (called FBF-f) we took $\gamma_n = \frac{1-\varepsilon}{1+\sqrt{8}}$ with $\varepsilon = \frac{1}{20(2+\sqrt{8})}$, while when h was evaluated via its proximal mapping (called FBF-b) we took $\gamma_n = \frac{1-\varepsilon}{\sqrt{8}}$ with $\varepsilon = \frac{1}{20(1+\sqrt{8})}$.

One can notice that from the point of view of the decrease of the objective function values generated by the two algorithms, the forward-backward-forward (FBF) primal-dual scheme proposed in [13] slightly overperforms the forward-backward (FB) one proposed in [20]. However, one should notice that (FBF) is more time and resources consuming then (FB), since for the first in each step the calculation of an additional iterate is needed.

Acknowledgements. The authors are grateful to Christopher Hendrich for the MATLAB implementations of the algorithms with respect to which the comparisons have been made and to an anonymous reviewer for comments and suggestions which improved the quality of the paper.

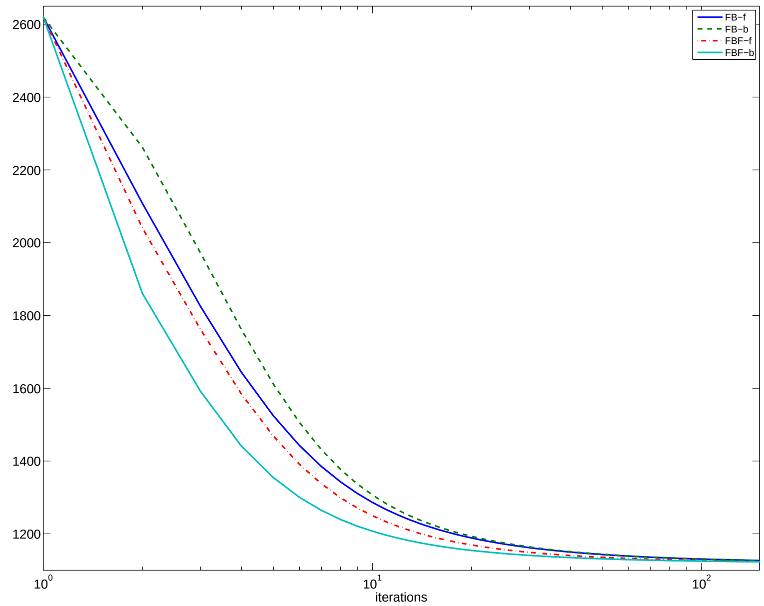

Figure 4: The figure shows the decrease of the objective function values for the averaged iterates generated by the (FB) and (FBF) algorithms.

References

[1] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011

[2] A. Beck, M. Teboulle, *A fast iterative shrinkage-tresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences 2(1), 183–202, 2009

[3] J.M. Borwein, J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, 2010

[4] R.I. Boţ, *Conjugate Duality in Convex Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer-Verlag, Berlin Heidelberg, 2010

[5] R.I. Boţ, E.R. Csetnek, *Regularity conditions via generalized interiority notions in convex optimization: new achievements and their relation to some classical statements*, Optimization 61(1), 35–65, 2012

[6] R.I. Boţ, E.R. Csetnek, A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximally monotone operators*, SIAM Journal on Optimization 23(4), 2011–2036, 2013

[7] R.I. Boţ, E.R. Csetnek, A. Heinrich, C. Hendrich, *On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems*, Mathematical Programming, DOI 10.1007/s10107-014-0766-0

[8] R.I. Boţ, C. Hendrich, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, Journal of Mathematical Imaging and Vision 49(3), 551–568, 2014

[9] L.M. Briceño-Arias, P.L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM Journal on Optimization 21(4), 1230–1250, 2011

[10] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision 20(1–2), 89–97, 2004

[11] A. Chambolle, T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision 40(1), 120–145, 2011

[12] P.L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization 53(5-6), 475–504, 2004

[13] P.L. Combettes, J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued and Variational Analysis 20(2), 307–330, 2012

[14] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications 158(2), 460–479, 2013

[15] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, 1976

[16] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, Dordrecht, 2004

[17] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14(5), 877–898, 1976

[18] S. Simons, *From Hahn-Banach to Monotonicity*, Springer-Verlag, Berlin, 2008

[19] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM Journal on Control and Optimization 38(2), 431–446, 2000

[20] B.C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics, 38(3), 667–681, 2013

[21] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002