A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators

Radu Ioan Boţ [*] Ernö Robert Csetnek [†] André Heinrich [‡]

July 15, 2013

Abstract. We consider the primal problem of finding the zeros of the sum of a maximal monotone operator and the composition of another maximal monotone operator with a linear continuous operator. By formulating its Attouch-Théra-type dual inclusion problem, a primal-dual splitting algorithm which simultaneously solves the two problems in finite-dimensional spaces is presented. The proposed scheme uses at each iteration the resolvents of the maximal monotone operators involved in separate steps and aims to overcome the shortcoming of classical splitting algorithms when dealing with compositions of maximal monotone and linear continuous operators. The iterative algorithm is used for solving nondifferentiable convex optimization problems arising in image processing and in location theory.

Key Words. maximal monotone operator, resolvent, operator splitting, subdifferential, minimization algorithm, duality

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

1 Introduction

1.1 Problem formulation and motivation

For X and Y real Hilbert spaces, $A : X \rightrightarrows X$ and $B : Y \rightrightarrows Y$ maximal monotone operators and $K : X \to Y$ a linear continuous operator we propose in this paper an iterative scheme for solving the monotone inclusion problem

$$\text{find } x \in X \text{ such that } 0 \in Ax + K^*BKx, \tag{1}$$

which makes separate use of the resolvents of A and B. The necessity of having such an algorithm is given by the fact that the classical splitting algorithms have considerable limitations when employed on the inclusion problem under investigation in its whole generality.

[*]University of Vienna, Faculty of Mathematics, Nordbergstraße 15, A-1090 Vienna, Austria, e-mail: radu.bot@univie.ac.at. Research partially supported by DFG (German Research Foundation), projects BO 2516/4-1 and WA922/1-3.

[†]University of Vienna, Faculty of Mathematics, Nordbergstraße 15, A-1090 Vienna, Austria, e-mail: ernoe.robert.csetnek@univie.ac.at. Research supported by DFG (German Research Foundation), project BO 2516/4-1.

[‡]Chemnitz University of Technology, Department of Mathematics, D-09107 Chemnitz, Germany, e-mail: andre.heinrich@mathematik.tu-chemnitz.de. Research supported by the European Union, the European Social Fund (ESF) and prudsys AG in Chemnitz.

Indeed, the Forward-Backward algorithm (see [5]) is a valuable option in this sense when B is single-valued and cocoercive, while the use of Tseng's Forward-Backward-Forward algorithm (see [28]) asks for B being single-valued and Lipschitz continuous on a superset of the image of the domain of A through K. On the other hand, the Douglas-Rachford algorithm (see [5,14]) asks for the maximal monotonicity of A and K^*BK and employs the resolvent of the latter, which can be expressed by means of the resolvent of B only in some very exceptional situations (see [5, Proposition 23.23]).

The aim of this article is to overcome this shortcoming by proposing a primal-dual splitting algorithm for simultaneously solving this monotone inclusion problem and its dual inclusion problem in the sense of Attouch-Théra (see [2,4,24])

$$\text{find } y \in Y \text{ such that } 0 \in B^{-1}y - KA^{-1}(-K^*)y. \tag{2}$$

In the formulation of the iterative scheme (only) simple forward evaluations of the operator K and of its adjoint are made, while the resolvents of A and B appear separately. More than that, the provided algorithm gives rise to a primal-dual iterative method for solving the monotone inclusion problem

$$\text{find } x \in X \text{ such that } 0 \in \sum_{i=1}^{k} K_i^* B_i K_i x, \tag{3}$$

and its dual problem

$$\text{find } y = (y_1, ..., y_k) \in Y_1 \times ... \times Y_k \text{ such that } \sum_{i=1}^{k} K_i^* y_i = 0 \text{ and } \bigcap_{i=1}^{k} (B_i K_i)^{-1}(y_i) \neq \emptyset, \tag{4}$$

where X and Y_i are real Hilbert spaces, $B_i : Y_i \rightrightarrows Y_i$ are maximal monotone operators and $K_i : X \to Y_i$ are linear continuous operators, which makes use of the resolvents of B_i and assumes only simple forward evaluations of $K_i, i = 1, ..., k$, (and of their adjoints).

Another primal-dual splitting algorithm which operates similarly with linear continuous operators has been recently introduced in [10,13]. By using a standard product space approach, this method basically reformulates the primal-dual pair as the problem of finding the zeros of the sum of a maximal monotone operator and a monotone and Lipschitz continuous operator, which is then solved by making use of the error-tolerant version of Tseng's Forward-Backward-Forward algorithm provided in [10].

When A and B are taken to be subdifferentials of proper, convex and lower semicontinuous functions, the primal-dual algorithm we propose in this paper becomes the iterative method from [12]. It has the advantage that, in the context of solving nonsmooth convex optimization problems involving compositions of proper, convex and lower semicontinuous functions with linear continuous operators, it only asks for the proximal points of the functions and not of the compositions. From this point of view, our approach proves to have a certain similariy with the algorithm proposed in [10,13]. In counterpart, popular decomposition algorithms like the augmented Lagrangian method (ALM) and the alternating direction method of multipliers (ADMM) (see [17] and, for some recent considerations on these methods, [9,16]) may assume the solving in each iteration of some optimization problems, for the solutions of which no explicit or closed form are always available.

The method we propose in this paper will open the gates towards proposing easy numerical implementations when solving regularized convex nondifferentiable problems of the shape

$$\inf_{x\in[0,1]^m}\left\{\|Ax-b\|^2+\lambda_1 TV(x)+\lambda_2\|x\|_1\right\},$$

where $TV:\mathbb{R}^m\to\mathbb{R}$ is a discrete total variation functional and $\lambda_1,\lambda_2>0$ are regularization parameters in the context of deblurring and denoising of images. Here, $A\in\mathbb{R}^{m\times m}$ describes a blur operator, $b\in\mathbb{R}^m$ represents the blurred and noisy image and the optimal solution is an estimation of the unknown original image, each of its pixels being assumed to range in the closed interval from 0 (pure black) to 1 (pure white). Beyond that, the primal-dual approach is expected to have a positive impact when dealing with the minimization of nonsmooth convex functions with an intricate formulation, as they occurr in applications in signal and video processing, machine learning, multifacilty location theory, portfolio optimization, average consensus on networks, etc.

The structure of the paper is the following. The remainder of this section is dedicated to some elements of convex analysis and of the theory of maximal monotone operators. In Section 2 we motivate and formulate a primal-dual splitting algorithm for solving the pair of monotone inclusion problems (1) - (2). Further, in Section 3 we address the primal-dual pair of monotone inclusion problems (3)-(4), while in Section 4 we employ the general schemes for several classes of nondifferentiable convex optimization problems. Finally, we consider applications of the presented algorithms in image deblurring and denoising and when solving the Fermat-Weber location problem and compare their performances to the ones of some iterative schemes recently introduced in the literature.

1.2 Monotone operators

In what follows we recall some elements of the theory of monotone operators in Hilbert spaces and refer the reader in this respect to [5, 26].

Let X be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|=\sqrt{\langle\cdot,\cdot\rangle}$. For an arbitrary set-valued operator $A:X\rightrightarrows X$ we denote by $\operatorname{Gr}A=\{(x,u)\in X\times X: u\in Ax\}$ its graph, by $\operatorname{dom}A=\{x\in X: Ax\neq\emptyset\}$ its domain and by $A^{-1}:X\rightrightarrows X$ its inverse operator, defined by $(u,x)\in\operatorname{Gr}A^{-1}$ if and only if $(x,u)\in\operatorname{Gr}A$. We say that A is monotone if $\langle x-y,u-v\rangle\geq 0$ for all $(x,u),(y,v)\in\operatorname{Gr}A$. A monotone operator A is said to be maximal monotone, if there exists no proper monotone extension of the graph of A on $X\times X$. A single-valued linear operator $A:X\to X$ is said to be skew, if $\langle x,Ax\rangle=0$ for all $x\in X$. Skew operators are maximal monotone.

The resolvent of A, $J_A:X\rightrightarrows X$, is defined by $J_A=(\operatorname{Id}_X+A)^{-1}$, where $\operatorname{Id}_X:X\to X,\operatorname{Id}_X(x)=x$ for all $x\in X$, is the identity operator on X. Moreover, if A is maximal monotone, then $J_A:X\to X$ is single-valued and maximal monotone (see [5, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma>0$ we have (see [5, Proposition 23.2])

$$p\in J_{\gamma A}x \text{ if and only if } (p,\gamma^{-1}(x-p))\in\operatorname{Gr}A$$

and (see [5, Proposition 23.18])

$$J_{\gamma A}+\gamma J_{\gamma^{-1}A^{-1}}\circ\gamma^{-1}\operatorname{Id}_X=\operatorname{Id}_X. \tag{5}$$

When Y is another Hilbert space and $K:X\to Y$ is a linear continuous operator, then $K^*:Y\to X$, defined by $\langle K^*y,x\rangle=\langle y,Kx\rangle$ for all $(x,y)\in X\times Y$, denotes the adjoint operator of K, while the norm of K is defined as $\|K\|=\sup\{\|Kx\|:x\in X,\|x\|\leq 1\}$.

2 A primal-dual splitting algorithm for finding the zeros of $A + K^*BK$

For X and Y real Hilbert spaces, $A : X \rightrightarrows X$ and $B : Y \rightrightarrows Y$ maximal monotone operators and $K : X \to Y$ a linear continuous operator we consider the problem of finding the pairs $(\widehat{x}, \widehat{y}) \in X \times Y$ fulfilling the system of inclusions

$$Kx \in B^{-1}y \text{ and } -K^*y \in Ax. \tag{6}$$

If $(\widehat{x}, \widehat{y})$ fulfills (6), then $\widehat{x}$ is a solution of the primal inclusion problem (1)

$$\text{find } x \in X \text{ such that } 0 \in Ax + K^*BKx$$

and $\widehat{y}$ is a solution of its dual inclusion problem (2)

$$\text{find } y \in Y \text{ such that } 0 \in B^{-1}y - KA^{-1}(-K^*)y.$$

On the other hand, if $\widehat{x} \in X$ is a solution of the problem (1), then there exists a solution $\widehat{y}$ of (2) such that $(\widehat{x}, \widehat{y})$ fulfills (6), while, if $\widehat{y} \in Y$ is a solution of the problem (2), then there exists a solution $\widehat{x}$ of (1) such that $(\widehat{x}, \widehat{y})$ fulfills (6). We refer the reader to [1,2,4,10,18,24] for more algorithmic and theoretical aspects concerning the primal-dual pair of inclusion problems (1)-(2).

For all $\sigma, \tau > 0$ it holds

$$\begin{aligned}(\widehat{x}, \widehat{y}) \text{ is a solution of (6)} &\Leftrightarrow \widehat{y} + \sigma K\widehat{x} \in (\mathrm{Id}_Y + \sigma B^{-1})\widehat{y} \text{ and } \widehat{x} - \tau K^*\widehat{y} \in (\mathrm{Id}_X + \tau A)\widehat{x} \\ &\Leftrightarrow \widehat{y} = J_{\sigma B^{-1}}(\widehat{y} + \sigma K\widehat{x}) \text{ and } \widehat{x} \in J_{\tau A}(\widehat{x} - \tau K^*\widehat{y}).\end{aligned} \tag{7}$$

The above equivalences motivate the following algorithm for solving (6).

Algorithm 1

`Initialization:` `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau\|K\|^2 < 1$ `and` $(x^0, y^0) \in X \times Y$.
`Set` $\overline{x}^0 := x^0$.

`For` $n \geq 0$ `set:`
$y^{n+1} := J_{\sigma B^{-1}}(y^n + \sigma K\overline{x}^n)$
$x^{n+1} := J_{\tau A}(x^n - \tau K^*y^{n+1})$
$\overline{x}^{n+1} := 2x^{n+1} - x^n$

Theorem 2 *Assume that the system of inclusions (6) has a solution $(\widehat{x}, \widehat{y}) \in X \times Y$ and let $(x^n, \overline{x}^n, y^n)_{n\geq 0}$ be the sequence generated by Algorithm 1. The following statements are true:*

(i) For any $n \geq 0$ it holds

$$\frac{\|x^n - \widehat{x}\|^2}{2\tau} + (1 - \sigma\tau\|K\|^2)\frac{\|y^n - \widehat{y}\|^2}{2\sigma} \leq \frac{\|x^0 - \widehat{x}\|^2}{2\tau} + \frac{\|y^0 - \widehat{y}\|^2}{2\sigma}, \tag{8}$$

thus the sequence $(x^n, y^n)_{n\geq 0}$ is bounded.

(ii) If X and Y are finite-dimensional, then the sequence $(x^n, y^n)_{n\geq 0}$ converges to a solution of the system of inclusions (6).

Proof. (i) The main idea of the proof relies on the following Fejér-type inequality

$$
\begin{aligned}
\frac{\|x^{n+1}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{n+1}-\widehat{y}\|^2}{2\sigma} \leq \frac{\|x^{n}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{n}-\widehat{y}\|^2}{2\sigma}+ \\
(-1+\sqrt{\sigma\tau}\|K\|)\frac{\|y^{n+1}-y^{n}\|^2}{2\sigma}-\frac{\|x^{n+1}-x^{n}\|^2}{2\tau}+\sqrt{\sigma\tau}\|K\|\frac{\|x^{n}-x^{n-1}\|^2}{2\tau} \\
-\langle K(x^{n+1}-x^n), y^{n+1}-\widehat{y}\rangle+\langle K(x^n-x^{n-1}), y^n-\widehat{y}\rangle,
\end{aligned}
\tag{9}
$$

which we will prove below to hold for any $n \geq 0$. By taking $N \in \mathbb{N}$ arbitrary such that $N \geq 2$ and by summing up the inequalities in (9) from $n=0$ to $N-1$ we obtain

$$
\begin{aligned}
\frac{\|x^{N}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{N}-\widehat{y}\|^2}{2\sigma} \leq \frac{\|x^{0}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{0}-\widehat{y}\|^2}{2\sigma}+ \\
(-1+\sqrt{\sigma\tau}\|K\|)\sum_{n=1}^{N}\frac{\|y^{n}-y^{n-1}\|^2}{2\sigma}+(-1+\sqrt{\sigma\tau}\|K\|)\sum_{n=1}^{N-1}\frac{\|x^{n}-x^{n-1}\|^2}{2\tau}- \\
\frac{\|x^{N}-x^{N-1}\|^2}{2\tau}-\langle K(x^N-x^{N-1}), y^N-\widehat{y}\rangle.
\end{aligned}
\tag{10}
$$

By combining (10) with

$$
-\langle K(x^N-x^{N-1}), y^N-\widehat{y}\rangle \leq \frac{\|x^{N}-x^{N-1}\|^2}{2\tau}+\frac{\tau\|K\|^2}{2}\|y^N-\widehat{y}\|^2,
$$

we get

$$
\begin{aligned}
\frac{\|x^{N}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{N}-\widehat{y}\|^2}{2\sigma} \leq \frac{\|x^{0}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{0}-\widehat{y}\|^2}{2\sigma}+ \\
(-1+\sqrt{\sigma\tau}\|K\|)\sum_{n=1}^{N}\frac{\|y^{n}-y^{n-1}\|^2}{2\sigma}+(-1+\sqrt{\sigma\tau}\|K\|)\sum_{n=1}^{N-1}\frac{\|x^{n}-x^{n-1}\|^2}{2\tau}+\frac{\tau\|K\|^2}{2}\|y^N-\widehat{y}\|^2
\end{aligned}
$$

or, equivalently,

$$
\begin{aligned}
\frac{\|x^{N}-\widehat{x}\|^2}{2\tau}+(1-\sigma\tau\|K\|^2)\frac{\|y^{N}-\widehat{y}\|^2}{2\sigma}+ \\
(1-\sqrt{\sigma\tau}\|K\|)\sum_{n=1}^{N}\frac{\|y^{n}-y^{n-1}\|^2}{2\sigma}+(1-\sqrt{\sigma\tau}\|K\|)\sum_{n=1}^{N-1}\frac{\|x^{n}-x^{n-1}\|^2}{2\tau} \leq \\
\frac{\|x^{0}-\widehat{x}\|^2}{2\tau}+\frac{\|y^{0}-\widehat{y}\|^2}{2\sigma}.
\end{aligned}
\tag{11}
$$

Since $\sigma\tau\|K\|^2<1$, (11) yields (8), hence $(x^n, y^n)_{n\geq 0}$ is bounded.

We are going now to prove that for any $n \geq 0$ the inequality (9) is fulfilled. For any $n \geq 0$ the iterations of Algorithm 1 yield that

$$
\left(x^{n+1}, \frac{1}{\tau}(x^n-\tau K^*y^{n+1}-x^{n+1})\right) \in \operatorname{Gr} A,
\tag{12}
$$

hence the monotonicity of A implies

$$
0 \leq \left\langle x^{n+1}-\widehat{x}, \frac{1}{\tau}(x^n-x^{n+1})-K^*y^{n+1}+K^*\widehat{y}\right\rangle.
\tag{13}
$$

Similarly, for any $n \geq 0$ we have

$$\left(y^{n+1}, \frac{1}{\sigma}(y^n + \sigma K\overline{x}^n - y^{n+1})\right) \in \operatorname{Gr} B^{-1}, \tag{14}$$

thus

$$0 \leq \left\langle K\overline{x}^n + \frac{1}{\sigma}(y^n - y^{n+1}) - K\widehat{x}, y^{n+1} - \widehat{y} \right\rangle. \tag{15}$$

On the other hand, for any $n \geq 0$ we have that

$$\|x^{n+1} - \widehat{x}\|^2 + \left\langle x^{n+1} - \widehat{x}, \frac{1}{\tau}(x^n - x^{n+1}) - K^*y^{n+1} + K^*\widehat{y} \right\rangle =$$

$$\left\langle x^{n+1} - \widehat{x}, \frac{1}{\tau}(x^n - \widehat{x}) + \left(1 - \frac{1}{\tau}\right)(x^{n+1} - \widehat{x}) - K^*y^{n+1} + K^*\widehat{y} \right\rangle =$$

$$\frac{1}{\tau}\langle x^{n+1} - \widehat{x}, x^n - \widehat{x}\rangle + \left(1 - \frac{1}{\tau}\right)\|x^{n+1} - \widehat{x}\|^2 + \langle K(x^{n+1} - \widehat{x}), -y^{n+1} + \widehat{y}\rangle,$$

hence

$$\frac{1}{\tau}\|x^{n+1} - \widehat{x}\|^2 + \left\langle x^{n+1} - \widehat{x}, \frac{1}{\tau}(x^n - x^{n+1}) - K^*y^{n+1} + K^*\widehat{y} \right\rangle =$$

$$\frac{1}{\tau}\langle x^{n+1} - \widehat{x}, x^n - \widehat{x}\rangle + \langle K(x^{n+1} - \widehat{x}), -y^{n+1} + \widehat{y}\rangle =$$

$$-\frac{1}{2\tau}\|x^{n+1} - x^n\|^2 + \frac{1}{2\tau}\|x^{n+1} - \widehat{x}\|^2 + \frac{1}{2\tau}\|\widehat{x} - x^n\|^2 + \langle K(x^{n+1} - \widehat{x}), -y^{n+1} + \widehat{y}\rangle,$$

where, for deriving the last formula, we use the identity

$$\langle a, b\rangle = -\frac{1}{2}\|a - b\|^2 + \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2 \ \forall a, b \in X.$$

Consequently, for any $n \geq 0$ it holds

$$\begin{aligned} &\frac{1}{2\tau}\|x^{n+1} - \widehat{x}\|^2 + \left\langle x^{n+1} - \widehat{x}, \frac{1}{\tau}(x^n - x^{n+1}) - K^*y^{n+1} + K^*\widehat{y} \right\rangle = \\ &-\frac{1}{2\tau}\|x^{n+1} - x^n\|^2 + \frac{1}{2\tau}\|x^n - \widehat{x}\|^2 + \langle K(x^{n+1} - \widehat{x}), -y^{n+1} + \widehat{y}\rangle. \end{aligned} \tag{16}$$

Thus, by combining (13) and (16), we get for any $n \geq 0$

$$\frac{1}{2\tau}\|x^{n+1} - \widehat{x}\|^2 \leq -\frac{1}{2\tau}\|x^{n+1} - x^n\|^2 + \frac{1}{2\tau}\|x^n - \widehat{x}\|^2 + \langle K(x^{n+1} - \widehat{x}), -y^{n+1} + \widehat{y}\rangle. \tag{17}$$

By proceeding in analogous manner we obtain the following estimate for any $n \geq 0$

$$\|y^{n+1} - \widehat{y}\|^2 + \left\langle K\overline{x}^n + \frac{1}{\sigma}(y^n - y^{n+1}) - K\widehat{x}, y^{n+1} - \widehat{y} \right\rangle =$$

$$\left\langle \frac{1}{\sigma}(y^n - \widehat{y}) + \left(1 - \frac{1}{\sigma}\right)(y^{n+1} - \widehat{y}) + K\overline{x}^n - K\widehat{x}, y^{n+1} - \widehat{y} \right\rangle =$$

$$\frac{1}{\sigma}\langle y^n - \widehat{y}, y^{n+1} - \widehat{y}\rangle + \left(1 - \frac{1}{\sigma}\right)\|y^{n+1} - \widehat{y}\|^2 + \langle K(\overline{x}^n - \widehat{x}), y^{n+1} - \widehat{y}\rangle,$$

hence

$$\frac{1}{\sigma}\|y^{n+1}-\widehat{y}\|^2+\left\langle K\overline{x}^n+\frac{1}{\sigma}(y^n-y^{n+1})-K\widehat{x},y^{n+1}-\widehat{y}\right\rangle =$$
$$\frac{1}{\sigma}\langle y^n-\widehat{y},y^{n+1}-\widehat{y}\rangle+\langle K(\overline{x}^n-\widehat{x}),y^{n+1}-\widehat{y}\rangle =$$
$$-\frac{1}{2\sigma}\|y^{n+1}-y^n\|^2+\frac{1}{2\sigma}\|y^{n+1}-\widehat{y}\|^2+\frac{1}{2\sigma}\|\widehat{y}-y^n\|^2+\langle K(\overline{x}^n-\widehat{x}),y^{n+1}-\widehat{y}\rangle.$$

From here we obtain for any $n\geq 0$

$$\begin{aligned}\frac{1}{2\sigma}\|y^{n+1}-\widehat{y}\|^2+\left\langle K\overline{x}^n+\frac{1}{\sigma}(y^n-y^{n+1})-K\widehat{x},y^{n+1}-\widehat{y}\right\rangle =\\ -\frac{1}{2\sigma}\|y^{n+1}-y^n\|^2+\frac{1}{2\sigma}\|y^n-\widehat{y}\|^2+\langle K(\overline{x}^n-\widehat{x}),y^{n+1}-\widehat{y}\rangle.\end{aligned} \tag{18}$$

and, thus, by combining (15) and (18), it follows

$$\frac{1}{2\sigma}\|y^{n+1}-\widehat{y}\|^2\leq -\frac{1}{2\sigma}\|y^{n+1}-y^n\|^2+\frac{1}{2\sigma}\|y^n-\widehat{y}\|^2+\langle K(\overline{x}^n-\widehat{x}),y^{n+1}-\widehat{y}\rangle. \tag{19}$$

Summing up the inequalities (17) and (19) and taking into account the definition of $\overline{x}^n$ we obtain for any $n\geq 0$

$$\begin{aligned}\frac{1}{2\tau}\|x^{n+1}-\widehat{x}\|^2+\frac{1}{2\sigma}\|y^{n+1}-\widehat{y}\|^2\leq\\ \frac{1}{2\tau}\|x^n-\widehat{x}\|^2+\frac{1}{2\sigma}\|y^n-\widehat{y}\|^2-\frac{1}{2\tau}\|x^{n+1}-x^n\|^2-\frac{1}{2\sigma}\|y^{n+1}-y^n\|^2+\\ \langle K(x^{n+1}+x^{n-1}-2x^n),-y^{n+1}+\widehat{y}\rangle,\end{aligned} \tag{20}$$

where $x^{-1}:=x^0$. Let us evaluate now the last term in relation (20). For any $n\geq 0$ it holds that

$$\begin{aligned}\langle K(x^{n+1}+x^{n-1}-2x^n),-y^{n+1}+\widehat{y}\rangle =\\ \langle K(x^{n+1}-x^n),-y^{n+1}+\widehat{y}\rangle+\langle K(x^n-x^{n-1}),y^n-\widehat{y}\rangle+\langle K(x^n-x^{n-1}),y^{n+1}-y^n\rangle\leq\\ -\langle K(x^{n+1}-x^n),y^{n+1}-\widehat{y}\rangle+\langle K(x^n-x^{n-1}),y^n-\widehat{y}\rangle+\|K\|\|x^n-x^{n-1}\|\|y^{n+1}-y^n\|\leq\\ -\langle K(x^{n+1}-x^n),y^{n+1}-\widehat{y}\rangle+\langle K(x^n-x^{n-1}),y^n-\widehat{y}\rangle+\\ \frac{\sqrt{\sigma\tau}\|K\|}{2\tau}\|x^n-x^{n-1}\|^2+\frac{\sqrt{\sigma\tau}\|K\|}{2\sigma}\|y^{n+1}-y^n\|^2.\end{aligned} \tag{21}$$

From (20) and (21) we obtain for any $n\geq 0$ the following estimation

$$\frac{1}{2\tau}\|x^{n+1}-\widehat{x}\|^2+\frac{1}{2\sigma}\|y^{n+1}-\widehat{y}\|^2\leq\frac{1}{2\tau}\|x^n-\widehat{x}\|^2+\frac{1}{2\sigma}\|y^n-\widehat{y}\|^2-$$
$$\frac{1}{2\tau}\|x^{n+1}-x^n\|^2-\frac{1}{2\sigma}\|y^{n+1}-y^n\|^2-\langle K(x^{n+1}-x^n),y^{n+1}-\widehat{y}\rangle+\langle K(x^n-x^{n-1}),y^n-\widehat{y}\rangle+$$
$$\frac{\sqrt{\sigma\tau}\|K\|}{2\tau}\|x^n-x^{n-1}\|^2+\frac{\sqrt{\sigma\tau}\|K\|}{2\sigma}\|y^{n+1}-y^n\|^2,$$

thus (9) holds. This concludes the proof of statement (i).

(ii) According to (i), $(x^n, y^n)_{n\geq 0}$ has a subsequence $(x^{n_k}, y^{n_k})_{k\geq 0}$ which converges to some element $(x^*, y^*) \in X \times Y$ as $k \to +\infty$. From (11) we also obtain that $\lim_{n\to+\infty}(x^n - x^{n-1}) = \lim_{n\to+\infty}(y^n - y^{n-1}) = 0$. Further, from (12) and (14) and using that, due to the maximal monotonicity of A and B, $\operatorname{Gr} A$ and $\operatorname{Gr} B$ are closed sets, it follows that (x^*, y^*) is a solution of the system of inclusions (6).

Further, let $k \geq 0$ and $N \in \mathbb{N}$, $N > n_k$. Summing up the inequalities in (9), for $(\widehat{x}, \widehat{y}) := (x^*, y^*)$, from $n = n_k$ to $N-1$ we obtain

$$\begin{aligned}
&\frac{\|x^N - x^*\|^2}{2\tau} + \frac{\|y^N - y^*\|^2}{2\sigma} + (1 - \sqrt{\sigma\tau}\|K\|) \sum_{n=n_k+1}^{N} \frac{\|y^n - y^{n-1}\|^2}{2\sigma} \\
&- \frac{\|x^{n_k} - x^{n_k-1}\|^2}{2\tau} + (1 - \sqrt{\sigma\tau}\|K\|) \sum_{n=n_k}^{N-1} \frac{\|x^n - x^{n-1}\|^2}{2\tau} + \frac{\|x^N - x^{N-1}\|^2}{2\tau} \\
&+\langle K(x^N - x^{N-1}), y^N - y^* \rangle - \langle K(x^{n_k} - x^{n_k-1}), y^{n_k} - y^* \rangle \\
&\leq \frac{\|x^{n_k} - x^*\|^2}{2\tau} + \frac{\|y^{n_k} - y^*\|^2}{2\sigma},
\end{aligned}$$

which yields

$$\begin{aligned}
\frac{\|x^N - x^*\|^2}{2\tau} + \frac{\|y^N - y^*\|^2}{2\sigma} &\leq \|K\| \|x^N - x^{N-1}\| \|y^N - y^*\| + \\
\frac{\|x^{n_k} - x^*\|^2}{2\tau} + \frac{\|y^{n_k} - y^*\|^2}{2\sigma} &+ \frac{\|x^{n_k} - x^{n_k-1}\|^2}{2\tau} + \langle K(x^{n_k} - x^{n_k-1}), y^{n_k} - y^* \rangle.
\end{aligned}$$

Consequently, by using also the boundedness of $(x^n, y^n)_{n\geq 0}$, for any $k \geq 0$ it holds

$$\limsup_{N\to+\infty} \left(\frac{\|x^N - x^*\|^2}{2\tau} + \frac{\|y^N - y^*\|^2}{2\sigma} \right) \leq$$

$$\frac{\|x^{n_k} - x^*\|^2}{2\tau} + \frac{\|y^{n_k} - y^*\|^2}{2\sigma} + \frac{\|x^{n_k} - x^{n_k-1}\|^2}{2\tau} + \langle K(x^{n_k} - x^{n_k-1}), y^{n_k} - y^* \rangle.$$

We finally let k converge to $+\infty$, which yields

$$\limsup_{N\to+\infty} \left(\frac{\|x^N - x^*\|^2}{2\tau} + \frac{\|y^N - y^*\|^2}{2\sigma} \right) = 0$$

and, further, $\lim_{N\to+\infty} x^N = x^*$ and $\lim_{N\to+\infty} y^N = y^*$. ■

We close this section by discussing another possible approach when solving the system of inclusions (6) by employing some ideas considered in [10, 13]. To this end we define the operators $M : X \times Y \rightrightarrows X \times Y$, $M(x, y) = (Ax, B^{-1}y)$, and $S : X \times Y \to X \times Y$, $S(x, y) = (K^*y, -Kx)$. The operator M is maximal monotone, since A and B are maximal monotone, while S is maximal monotone, since it is a skew linear operator. Then $(\widehat{x}, \widehat{y}) \in X \times Y$ is a solution of the system of inclusions (6) if and only if it solves the inclusion problem

$$\text{find } (x, y) \in X \times Y \text{ such that } (0, 0) \in S(x, y) + M(x, y). \tag{22}$$

Applying Algorithm 1 to the problem (22) with starting point $(x^0, y^0, u^0, v^0) \in X \times Y \times X \times Y$, $(\overline{x}^0, \overline{y}^0) = (x^0, y^0)$ and $\sigma, \tau > 0$ gives rise for any $n \geq 0$ to the following iterations:

$$\begin{aligned}
(u^{n+1}, v^{n+1}) &:= J_{\sigma M^{-1}}\big[(u^n, v^n) + \sigma(\overline{x}^n, \overline{y}^n)\big] \\
(x^{n+1}, y^{n+1}) &:= J_{\tau S}\big[(x^n, y^n) - \tau(u^{n+1}, v^{n+1})\big] \\
(\overline{x}^{n+1}, \overline{y}^{n+1}) &:= 2(x^{n+1}, y^{n+1}) - (x^n, y^n).
\end{aligned}$$

Since

$$J_{\sigma M^{-1}} = J_{\sigma A^{-1}} \times J_{\sigma B}$$

and (see [10, Proposition 2.7])

$$J_{\tau S}(x, y) = \big((\mathrm{Id}_X + \tau^2 K^* K)^{-1}(x - \tau K^* y), (\mathrm{Id}_Y + \tau^2 K K^*)^{-1}(y + \tau K x)\big) \ \forall (x, y) \in X \times Y,$$

this yields the following algorithm:

Algorithm 3

`Initialization`: `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau < 1$ `and` $(x^0, y^0), (u^0, v^0) \in X \times Y$.
`Set` $(\overline{x}^0, \overline{y}^0) := (x^0, y^0)$.

`For` $n \geq 0$ `set`:

$$\begin{aligned}
u^{n+1} &:= J_{\sigma A^{-1}}(u^n + \sigma \overline{x}^n) \\
v^{n+1} &:= J_{\sigma B}(v^n + \sigma \overline{y}^n) \\
x^{n+1} &:= (\mathrm{Id}_X + \tau^2 K^* K)^{-1}\big[x^n - \tau u^{n+1} - \tau K^*(y^n - \tau v^{n+1})\big] \\
y^{n+1} &:= (\mathrm{Id}_Y + \tau^2 K K^*)^{-1}\big[y^n - \tau v^{n+1} + \tau K(x^n - \tau u^{n+1})\big] \\
\overline{x}^{n+1} &:= 2x^{n+1} - x^n \\
\overline{y}^{n+1} &:= 2y^{n+1} - y^n
\end{aligned}$$

The following convergence statement is a consequence of Theorem 2.

Theorem 4 *Assume that X and Y are finite-dimensional spaces and that the system of inclusions* (6) *is solvable. Then the sequence $(x^n, y^n)_{n \geq 0}$ generated in Algorithm 3 converges to (x^*, y^*), a solution of the system of inclusions* (6)*, which yields that x^* is a solution of the primal inclusion problem* (1) *and y^* is a solution of the dual inclusion problem* (2).

Remark 5 As we have already mentioned, the system of inclusions (6) is solvable if and only if the primal inclusion problem (1) is solvable, which is further equivalent to solvability of the dual inclusion problem (2). Let us also notice that from the point of view of the numerical implementation Algorithm 3 has the drawback to ask for the calculation of the inverses of $\mathrm{Id}_X + \tau^2 K^* K$ and $\mathrm{Id}_Y + \tau^2 K K^*$. This task can be in general very hard, but it becomes very simple when K is, for instance, orthogonal, like it happens for the linear transformations to which orthogonal wavelets give rise and which play an important role in signal processing.

3 Zeros of sums of compositions of monotone operators with linear continuous operators

In this section we provide via the primal-dual scheme Algorithm 1 an algorithm for solving the inclusion problem (3)

$$\text{find } x \in X \text{ such that } 0 \in \sum_{i=1}^{k} K_i^* B_i K_i x,$$

where X and Y_i are real Hilbert spaces, $B_i : Y_i \rightrightarrows Y_i$ are maximal monotone operators and $K_i : X \to Y_i$ are linear continuous operators for $i = 1, ..., k$. The dual inclusion problem of (3) is problem (4)

$$\text{find } y = (y_1, ..., y_k) \in Y_1 \times ... \times Y_k \text{ such that } \sum_{i=1}^{k} K_i^* y_i = 0 \text{ and } \bigcap_{i=1}^{k} (B_i K_i)^{-1}(y_i) \neq \emptyset.$$

Following the product space approach from [10] and [5] we show that this primal-dual pair can be reduced to a primal-dual pair of inclusion problems of the form (1)-(2).

Consider the real Hilbert space $H := X^k$ endowed with the inner product $\langle x, u\rangle_H = \sum_{i=1}^{k} \langle x_i, u_i\rangle_X$ for $x = (x_i)_{1\le i\le k}, u = (u_i)_{1\le i\le k} \in H$, where $\langle\cdot,\cdot\rangle_X$ denotes the inner product on X. Further, let $Y := Y_1 \times ... \times Y_k$ be the real Hilbert space endowed with the inner product $\langle y, z\rangle_Y := \sum_{i=1}^{k} \langle y_i, z_i\rangle_{Y_i}$ for $y = (y_i)_{1\le i\le k}, z = (z_i)_{1\le i\le k} \in Y$, where $\langle\cdot,\cdot\rangle_{Y_i}$ denotes the inner product on $Y_i, i = 1, ..., k$. We define $A : H \rightrightarrows H$, $A := N_V$, where $V = \{(x, ..., x) \in H : x \in X\}$, $B : Y \rightrightarrows Y$, $B(y_1, ..., y_k) = (B_1 y_1, ..., B_k y_k)$, and $K : H \to Y$, $K(x_1, ..., x_k) = (K_1 x_1, ..., K_k x_k)$. Obviously, the adjoint operator of K is $K^* : Y \to H$, $K^*(y_1, ..., y_k) = (K_1^* y_1, ..., K_k^* y_k)$, for $(y_1, ..., y_k) \in Y$. Further, let be $j : X \to H$, $j(x) = (x, ..., x)$.

The operators A and B are maximal monotone and

$$x \text{ solves (3) if and only if } (0, ..., 0) \in A(j(x)) + K^*BK(j(x)),$$

while

$$y = (y_1, ..., y_k) \text{ solves (4) if and only if } (0, ..., 0) \in B^{-1}y - KA^{-1}(-K^*)y.$$

Applying Algorithm 1 to the inclusion problem

$$\text{find } (x_1, ..., x_k) \in H \text{ such that } 0 \in A(x_1, ..., x_k) + K^*BK(x_1, ..., x_k) \tag{23}$$

with starting point $(x^0, ..., x^0, y_1^0, ..., y_k^0) \in \underbrace{X \times ... \times X}_{k} \times Y_1 \times ... \times Y_k$, constants $\sigma, \tau > 0$ and $(\overline{x}_1^0, ..., \overline{x}_k^0) := (x^0, ..., x^0)$ yields for any $n \ge 0$ the following iterations:

$$\begin{aligned}
(y_i^{n+1})_{1\le i\le k} &:= J_{\sigma B^{-1}}\Big((y_i^n)_{1\le i\le k} + \sigma K(\overline{x}_i^n)_{1\le i\le k}\Big)\\
(x_i^{n+1})_{1\le i\le k} &:= J_{\tau A}\Big((x_i^n)_{1\le i\le k} - \tau K^*(y_i^{n+1})_{1\le i\le k}\Big)\\
(\overline{x}_i^{n+1})_{1\le i\le k} &:= 2(x_i^{n+1})_{1\le i\le k} - (x_i^n)_{1\le i\le k}.
\end{aligned}$$

According to [10], for the occurring resolvents we have that $J_{\tau A}(u_1, ..., u_k) = j(\frac{1}{k}\sum_{i=1}^{k} u_i)$ for $(u_1, ..., u_k) \in H$ and $J_{\sigma B^{-1}}(z_1, ..., z_k) = (J_{\sigma B_1^{-1}} z_1, ..., J_{\sigma B_k^{-1}} z_k)$ for $(z_1, ..., z_k) \in Y$. This means that for any $n \ge 1$ it holds $x_1^n = ... = x_k^n$ and $\overline{x}_1^{n+1} = ... = \overline{x}_k^{n+1}$, which shows that there is no loss in the generality of the algorithm when assuming that the first k components of the starting point coincide. Notice that a solution $(\widehat{x}_1, ..., \widehat{x}_k)$ of (23) must belong to $\operatorname{dom} A$, thus $\widehat{x}_1 = ... = \widehat{x}_k$. We obtain the following algorithm:

Algorithm 6

`Initialization`*:* `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau\sum_{i=1}^{k} \|K_i\|^2 < 1$ `and` $(x^0, y_1^0, ..., y_k^0) \in X \times Y_1 \times ... \times Y_k$. `Set` $\overline{x}^0 := x^0$.

`For` $n \ge 0$ `set`*:*
$y_i^{n+1} := J_{\sigma B_i^{-1}}(y_i^n + \sigma K_i \overline{x}^n), i = 1, ..., k$
$x^{n+1} := x^n - \frac{\tau}{k}\sum_{i=1}^{k} K_i^* y_i^{n+1}$
$\overline{x}^{n+1} := 2x^{n+1} - x^n$

The convergence of Algorithm 6 is stated by the following result which is a consequence of Theorem 2.

Theorem 7 *Assume that X and $Y_i, i = 1, ..., k$, are finite-dimensional spaces and (3) is solvable. Then (4) is also solvable and the sequences $(x^n)_{n\geq 0}$ and $(y_1^n, ..., y_k^n)_{n\geq 0}$ generated in Algorithm 6 converge to a solution of the primal inclusion problem (3) and to a solution of the dual inclusion problem (4), respectively.*

Remark 8 Since $\|K\|^2 \leq \sum_{i=1}^k \|K_i\|^2$, the inequality $\sigma\tau \sum_{i=1}^k \|K_i\|^2 < 1$ in Algorithm 6 is considered in order to ensure that $\sigma\tau\|K\|^2 < 1$.

Remark 9 The iterative schemes in Algorithm 1 and Algorithm 6 follow for some appropriate initializations from the one proposed in [29], as the author of this paper kindly pointed it out to us. The approach in [29] relies on an idea similar to the one from [10,13], as it assumes the reformulation of the primal-dual pair as the problem of finding the zeros of the sum of a maximal monotone operator and a cocoercive operator, solved via an error-tolerant Forward-Backward algorithm. In counterpart, the technique we propose here puts in foreground the Fejér-type inequality (9), which is of big importance in the context of studying the convergence rates of iterative methods relying on Algorithm 1.

When particularizing the above framework to the case when $Y_i = X$ and $K_i = \mathrm{Id}_X$ for $i = 1, .., k$, the primal-dual pair of inclusion problems (3)-(4) become

$$\text{find } x \in X \text{ such that } 0 \in \sum_{i=1}^k B_i x \tag{24}$$

and

$$\text{find } y = (y_1, ..., y_k) \in X \times ... \times X \text{ such that } \sum_{i=1}^k y_i = 0 \text{ and } \bigcap_{i=1}^k B_i^{-1}(y_i) \neq \emptyset, \tag{25}$$

respectively. In this situation $H = Y$, $K = \mathrm{Id}_H$, $\|K\| = 1$ and

$$x \text{ solves (24) if and only if } (0, ..., 0) \in A(j(x)) + B(j(x)),$$

while

$$y = (y_1, ..., y_k) \text{ solves (25) if and only if } (0, ..., 0) \in B^{-1}(y) - A^{-1}(-y).$$

Algorithm 6 yields in this particular case the following iterative scheme:

Algorithm 10

`Initialization`*:* `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau < 1$ `and` $(x^0, y_1^0, ..., y_k^0) \in \underbrace{X \times ... \times X}_{k+1}$. `Set` $\overline{x}^0 := x^0$.

`For` $n \geq 0$ `set`*:*
$$y_i^{n+1} := J_{\sigma B_i^{-1}}(y_i^n + \sigma\overline{x}^n), i = 1, ..., k$$
$$x^{n+1} := x^n - \tfrac{\tau}{k}\textstyle\sum_{i=1}^k y_i^{n+1}$$
$$\overline{x}^{n+1} := 2x^{n+1} - x^n$$

The convergence of Algorithm 10 follows via Theorem 7.

Theorem 11 *Assume that X is a finite-dimensional space and* (24) *is solvable. Then* (25) *is also solvable and the sequences $(x^n)_{n\geq 0}$ and $(y_1^n, ..., y_k^n)_{n\geq 0}$ generated in Algorithm 10 converge to a solution of the primal inclusion problem* (24) *and to a solution of the dual inclusion problem* (25)*, respectively.*

In the last part of this section we provide a second algorithm which solves (24) and (25) which starts from the premise that by changing the roles of A and B one has

$$x \text{ solves (24) if and only if } (0, ..., 0) \in B(j(x)) + A(j(x)),$$

while

$$y = (y_1, ..., y_k) \text{ solves (25) if and only if } (0, ..., 0) \in A^{-1}(-y) - B^{-1}(y).$$

Applying Algorithm 1 to the inclusion problem

$$\text{find } (x_1, ..., x_k) \in H \text{ such that } 0 \in B(x_1, ..., x_k) + A(x_1, ..., x_k)$$

with starting point $(x_1^0, ..., x_k^0, y_1^0, ..., y_k^0) \in \underbrace{X \times ... \times X}_{2k}$, constants $\sigma, \tau > 0$ and $(\overline{x}_1^0, ..., \overline{x}_k^0) :=$ $(x_1^0, ..., x_k^0)$ yields for any $n \geq 0$ the following iterations:

$$\begin{aligned}
(y_i^{n+1})_{1\leq i\leq k} &= J_{\sigma A^{-1}}\Big((y_i^n)_{1\leq i\leq k} + \sigma(\overline{x}_i^n)_{1\leq i\leq k}\Big)\\
(x_i^{n+1})_{1\leq i\leq k} &= J_{\tau B}\Big((x_i^n)_{1\leq i\leq k} - \tau(y_i^{n+1})_{1\leq i\leq k}\Big)\\
(\overline{x}_i^{n+1})_{1\leq i\leq k} &= 2(x_i^{n+1})_{1\leq i\leq k} - (x_i^n)_{1\leq i\leq k}.
\end{aligned}$$

Noticing that $J_{\sigma A^{-1}} = J_{\sigma N_V^{-1}} = \mathrm{Id}_H - \sigma J_{\sigma^{-1} N_V} \circ \sigma^{-1}\mathrm{Id}_H$ (see [5, Proposition 23.18]) and $J_{\sigma^{-1} N_V}(u_1, ..., u_k) = J_{N_V}(u_1, ..., u_k) = j(\frac{1}{k}\sum_{i=1}^k u_i)$ for $(u_1, ..., u_k) \in H$ (see [10, relation (3.27)]) and by making for any $n \geq 0$ the change of variables $y_i^n := -y_i^n$ for $i = 1, ..., k$, we obtain the following iterative scheme:

Algorithm 12

`Initialization`*:*	`Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau < 1$ `and` $(x_1^0, ..., x_k^0, y_1^0, ..., y_k^0) \in \underbrace{X \times ... \times X}_{2k}$. `Set` $(\overline{x}_1^0, ..., \overline{x}_k^0) := (x_1^0, ..., x_k^0)$.
`For` $n \geq 0$ `set`*:*	$y_i^{n+1} := y_i^n - \sigma\overline{x}_i^n + \frac{1}{k}\sum_{j=1}^k y_j^n + \frac{\sigma}{k}\sum_{j=1}^k \overline{x}_j^n, i = 1, ..., k$ $x_i^{n+1} := J_{\tau B_i}(x_i^n + \tau y_i^{n+1}), i = 1, ..., k$ $\overline{x}_i^{n+1} := 2x_i^{n+1} - x_i^n, i = 1, ..., k$

Theorem 13 *Assume that X is finite dimensional and* (24) *is solvable. Then* (25) *is also solvable, the sequences $(x_i^n)_{n\geq 0}, i = 1, ..., k$, generated in Algorithm 12 converge to the same solution of* (24) *and the sequence $(y_1^n, ..., y_k^n)_{n\geq 0}$ generated by the same algorithm converges to a solution of* (25)*.*

4 Solving convex optimization problems via the primal-dual algorithm

The aim of this section is to employ the iterative methods investigated above for solving several classes of unconstrained convex optimization problems. We start by recalling some elements of convex analysis (see [5, 6, 19, 26, 31]).

When X is a real Hilbert space and $f : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ a proper (i.e., $\operatorname{dom} f = \{x \in X : f(x) < +\infty\} \neq \emptyset$ and $f > -\infty$), convex and lower semicontinuous function $f : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, then its (convex) subdifferential, which is defined by $\partial f(x) := \{u \in X : f(y) \geq f(x) + \langle u, y - x\rangle \ \forall y \in X\}$, for $x \in f^{-1}(\mathbb{R})$, and $\partial f(x) := \emptyset$, otherwise, is a maximal monotone operator (see [25]). Moreover, $(\partial f)^{-1} = \partial f^*$, where $f^* : X \to \overline{\mathbb{R}}$, $f^*(u) = \sup_{x \in X}\{\langle u, x\rangle - f(x)\}$ for $u \in X$, denotes the conjugate function of f. Examples of maximal monotone operators which fail to be subdifferentials of a proper, convex and lower semicontinuous function are nonzero skew operators defined on a Hilbert space with a dimension greater than or equal to 2 (see [26]).

For $\gamma > 0$ and $x \in X$ 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

$$\inf_{y \in X} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{26}$$

Notice that $J_{\gamma\partial f} = (\operatorname{Id}_X + \gamma\partial f)^{-1} = \operatorname{prox}_{\gamma f}$, thus $\operatorname{prox}_{\gamma f} : X \to X$ is a single-valued operator fulfilling the extended Moreau's decomposition formula

$$\operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{(1/\gamma) f^*} \circ \gamma^{-1} \operatorname{Id}_X = \operatorname{Id}_X. \tag{27}$$

Let us also recall that the function $f : X \to \overline{\mathbb{R}}$ is said to be strongly convex (with modulus $\gamma > 0$), if $f - \frac{\gamma}{2}\|\cdot\|^2$ is a convex function.

When $C \subseteq X$ is a nonempty, convex and closed set, its indicator function $\delta_C : X \to \overline{\mathbb{R}}$, is the proper, convex and lower semicontinuous function that takes the value 0 on C and $+\infty$ otherwise. The subdifferential of δ_C is the normal cone of C, that is $N_C(x) = \{u \in X : \langle u, y - x\rangle \leq 0 \ \forall y \in C\}$, if $x \in C$ and $N_C(x) = \emptyset$ for $x \notin C$. For $\gamma > 0$ it holds

$$J_{\gamma N_C} = J_{N_C} = J_{\partial\delta_C} = (\operatorname{Id}_X + N_C)^{-1} = \operatorname{prox}_{\delta_C} = P_C, \tag{28}$$

where $P_C : X \to C$ denotes the projection operator on C (see [5, Example 23.3 and Example 23.4]).

For the real Hilbert spaces X and Y, the proper, convex and lower semicontinuous functions $f : X \to \overline{\mathbb{R}}$ and $g : Y \to \overline{\mathbb{R}}$ and the linear continuous operator $K : X \to Y$ consider the optimization problem

$$\inf_{x \in X}\{f(x) + g(Kx)\} \tag{29}$$

along with its Fenchel dual problem (see [5, 6, 19, 31])

$$\sup_{y \in Y}\{-f^*(-K^*y) - g^*(y)\}. \tag{30}$$

For this primal-dual pair weak duality always holds, i.e., the optimal objective value of the primal problem is greater than or equal to the optimal objective value of the dual problem.

In order to guarantee strong duality, i.e., the situation when the optimal objective values of the two problems coincide and the dual problem has an optimal solution one needs to ask for the fulfillment of a so-called qualification condition. One of the weakest interiority-type qualification conditions known in the literature reads (see [5–7, 26, 31])

$$(QC) \quad \Big| \quad 0 \in \operatorname{sqri}(\operatorname{dom} g - K(\operatorname{dom} f)).$$

Here, for C a convex set, we denote by $\operatorname{sqri} C := \{x \in C : \cup_{\lambda > 0} \lambda(C - x)$ is a closed linear subspace of $X\}$ its strong quasi-relative interior. The strong quasi-relative interior of C is a superset of the topological interior of C, i.e., $\operatorname{int} C \subseteq \operatorname{sqri} C$ (in general this inclusion may be strict). If X is finite-dimensional, then $\operatorname{sqri} C$ coincides with the relative interior of C, which is the interior of C with respect to its affine hull. The qualification condition (QC) is fulfilled, for instance, when there exists $x' \in \operatorname{dom} f \cap K^{-1}(\operatorname{dom} g)$ such that g is continuous at Kx'.

Algorithm 1 written for $A := \partial f$ and $B := \partial g$ yields the following iterative scheme:

Algorithm 14

`Initialization`*:* `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau\|K\|^2 < 1$ `and` $(x^0, y^0) \in X \times Y$.
`Set` $\overline{x}^0 := x^0$.

`For` $n \geq 0$ `set`*:*
$$y^{n+1} := \operatorname{prox}_{\sigma g^*}(y^n + \sigma K\overline{x}^n)$$
$$x^{n+1} := \operatorname{prox}_{\tau f}(x^n - \tau K^* y^{n+1})$$
$$\overline{x}^{n+1} := 2x^{n+1} - x^n$$

We have the following convergence result.

Theorem 15 *Assume that the primal problem* (29) *has an optimal solution* $\widehat{x}$ *and the qualification condition* (QC) *is fulfilled. Let* $(x^n, \overline{x}^n, y^n)_{n \geq 0}$ *be the sequence generated by Algorithm 14. The following statements are true:*

(i) There exists $\widehat{y} \in Y$*, an optimal solution of the dual problem* (30)*, the optimal objective values of the two optimization problems coincide and* $(\widehat{x}, \widehat{y})$ *is a solution of the system of inclusions*

$$Kx \in \partial g^*(y) \text{ and } -K^* y \in \partial f(x). \tag{31}$$

(ii) For any $n \geq 0$ *it holds*

$$\frac{\|x^n - \widehat{x}\|^2}{2\tau} + (1 - \sigma\tau\|K\|^2)\frac{\|y^n - \widehat{y}\|^2}{2\sigma} \leq \frac{\|x^0 - \widehat{x}\|^2}{2\tau} + \frac{\|y^0 - \widehat{y}\|^2}{2\sigma}, \tag{32}$$

thus the sequence $(x^n, y^n)_{n \geq 0}$ *is bounded.*

(iii) If X *and* Y *are finite-dimensional, then* $(x^n)_{n \geq 0}$ *converges to an optimal solution of* (29) *and* $(y^n)_{n \geq 0}$ *converges to an optimal solution of* (30).

Remark 16 (i) Statement (i) of the above theorem is well-known in the literature, (31) being nothing else than the system of optimality conditions for the primal-dual pair (29)-(30) (see [6, 19, 31]), while the other two statements follow from Theorem 2.

(ii) The existence of optimal solutions of the primal problem (29) is guaranteed if $K(\operatorname{dom} f) \cap \operatorname{dom} g \neq \emptyset$ and, for instance, f is coercive and g is bounded below. Indeed, under these circumstances, the objective function of (29) is coercive and the statement follows via [31, Theorem 2.5.1(ii)]. On the other hand, when f is strongly convex, then

$f+g\circ K$ is strongly convex, too, thus (29) has a unique optimal solution (see [5, Corollary 11.16]).

(iii) We rediscovered above the iterative scheme and the convergence statement from [12] as a particular instance of the general results furnished in Section 2.

For X and Y_i real Hilbert spaces, $g_i : Y_i \rightarrow \overline{\mathbb{R}}$ proper, convex and lower semicontinuous functions and $K_i : X \rightarrow Y_i$ linear continuous operators, $i = 1,...,k$, consider the optimization problem

$$\inf_{x\in X} \sum_{i=1}^{k} g_i(K_i x) \tag{33}$$

and its Fenchel-type dual problem

$$\sup_{\substack{y_i\in Y_i, i=1,...,k \\ \sum_{i=1}^k K_i^* y_i = 0}} \sum_{i=1}^{k} -g_i^*(y_i). \tag{34}$$

One of the weakest qualification conditions guaranteeing strong-duality for the primal-dual pair (33)-(34) reads (see, for instance, [6, 10, 31])

$$(QC^{\Sigma}) \quad \Big| \quad 0 \in \operatorname{sqri}\left(\prod_{i=1}^k \operatorname{dom} g_i - \{(K_1x,...,K_kx) : x \in X\}\right)$$

and it is fulfilled, for instance, when there exists $x' \in \bigcap_{i=1}^k K_i^{-1}(\operatorname{dom} g_i)$ such that g_i is continuous at $K_i x'$, $i = 1,...,k$. By taking $B_i := \partial g_i$, $i = 1,...,k$, Algorithm 6 yields the following iterative scheme:

Algorithm 17

`Initialization`*:*	`Choose` $\sigma,\tau > 0$ `such that` $\sigma\tau\sum_{i=1}^k \|K_i\|^2 < 1$ `and` $(x^0, y_1^0, ..., y_k^0) \in X \times Y_1 \times ... \times Y_k$. `Set` $\overline{x}^0 := x^0$.
`For` $n \geq 0$ `set`*:*	$y_i^{n+1} := \operatorname{prox}_{\sigma g_i^*}(y_i^n + \sigma K_i \overline{x}^n), i = 1, ..., k$ $x^{n+1} := x^n - \frac{\tau}{k}\sum_{i=1}^k K_i^* y_i^{n+1}$ $\overline{x}^{n+1} := 2x^{n+1} - x^n$

The convergence of Algorithm 17 is stated by the following result which is a consequence of Theorem 7.

Theorem 18 *Assume that the primal problem* (33) *has an optimal solution* $\widehat{x}$ *and the qualification conditions* (QC^{Σ}) *is fulfilled. The following statements are true:*

(i) There exists $(\widehat{y}_1, ..., \widehat{y}_k) \in Y_1 \times ... \times Y_k$*, an optimal solution of the dual problem* (34)*, the optimal objective values of the two optimization problems coincide and* $(\widehat{x}, \widehat{y}_1, ..., \widehat{y}_k)$ *is a solution of the system of inclusions*

$$K_i x \in \partial g_i^*(y_i), i = 1, ..., k, \text{ and } \sum_{i=1}^{k} K_i^* y_i = 0. \tag{35}$$

(ii) If X *and* Y *are finite-dimensional, then the sequences* $(x^n)_{n\geq 0}$ *and* $(y_1^n, ..., y_k^n)_{n\geq 0}$ *generated in Algorithm 17 converge to an optimal solution of* (33) *and* (34)*, respectively.*

Considering, finally, the particular case when $Y_i = X$ and $K_i = \mathrm{Id}_X$, $i = 1, ..., k$, the problems (33) and (34) become

$$\inf_{x \in X} \sum_{i=1}^{k} g_i(x) \tag{36}$$

and, respectively,

$$\sup_{\substack{y_i \in X, i=1,...,k \\ \sum_{i=1}^{k} y_i = 0}} \sum_{i=1}^{k} -g_i^*(y_i). \tag{37}$$

The qualification conditions (QC^{Σ}) looks in this case like

$$(QC^{id}) \quad \Big| \quad 0 \in \operatorname{sqri}\left(\prod_{i=1}^{k} \operatorname{dom} g_i - \{(x, ..., x) : x \in X\}\right)$$

and it is fulfilled, for instance, if there exists $x' \in \cap_{i=1}^{k} \operatorname{dom} g_i$ such that $k-1$ of the functions $g_i, i = 1, ..., k$, are continuous at x'. By particularizing Algorithm 17 we obtain:

Algorithm 19

`Initialization`: `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau < 1$ `and` $(x^0, y_1^0, ..., y_k^0) \in \underbrace{X \times ... \times X}_{k+1}$. `Set` $\overline{x}^0 := x^0$.

`For` $n \geq 0$ `set`:
$$\begin{aligned} y_i^{n+1} &:= \operatorname{prox}_{\sigma g_i^*}(y_i^n + \sigma \overline{x}^n), i = 1, ..., k \\ x^{n+1} &:= x^n - \tfrac{\tau}{k} \textstyle\sum_{i=1}^{k} y_i^{n+1} \\ \overline{x}^{n+1} &:= 2x^{n+1} - x^n \end{aligned}$$

while Algorithm 12 gives rise to the following iterative scheme:

Algorithm 20

`Initialization`: `Choose` $\sigma, \tau > 0$ `such that` $\sigma\tau < 1$ `and` $(x_1^0, ..., x_k^0, y_1^0, ..., y_k^0) \in \underbrace{X \times ... \times X}_{2k}$. `Set` $(\overline{x}_1^0, ..., \overline{x}_k^0) := (x_1^0, ..., x_k^0)$.

`For` $n \geq 0$ `set`:
$$\begin{aligned} y_i^{n+1} &:= y_i^n - \sigma \overline{x}_i^n + \tfrac{1}{k} \textstyle\sum_{j=1}^{k} y_j^n + \tfrac{\sigma}{k} \sum_{j=1}^{k} \overline{x}_j^n, i = 1, ..., k \\ x_i^{n+1} &:= \operatorname{prox}_{\tau g_i}(x_i^n + \tau y_i^{n+1}), i = 1, ..., k \\ \overline{x}_i^{n+1} &:= 2x_i^{n+1} - x_i^n, i = 1, ..., k \end{aligned}$$

We have the following convergence theorem.

Theorem 21 *Assume that the primal problem* (36) *has an optimal solution* $\widehat{x}$ *and the qualification conditions* (QC^{id}) *is fulfilled. The following statements are true:*

(i) There exists $(\widehat{y}_1, ..., \widehat{y}_k) \in X \times ... \times X$*, an optimal solution of the dual problem* (37)*, the optimal objective values of the two optimization problems coincide and* $(\widehat{x}, \widehat{y}_1, ..., \widehat{y}_k)$ *is a solution of the system of inclusions*

$$x \in \partial g_i^*(y_i), i = 1, ..., k, \text{ and } \sum_{i=1}^{k} y_i = 0. \tag{38}$$

(ii) If X *is finite-dimensional, then the sequences* $(x^n)_{n \geq 0}$ *and* $(y_1^n, ..., y_k^n)_{n \geq 0}$ *generated in Algorithm 19 converge to an optimal solution of* (36) *and* (37)*, respectively.*

(iii) If X *is finite-dimensional, then the sequences* $(x_i^n)_{n \geq 0}, i = 1, ..., k$*, generated in Algorithm 20 converges to the same optimal solution of* (36) *and the sequence* $(y_1^n, ...y_k^n)_{n \geq 0}$ *generated by the same algorithm converges to an optimal solution of* (37)*.*

Due to the absence of linear continuous operators in its formulation, one can consider for the solving of the optimization problem (36) also other splitting algorithms with an easy implementable formulation. In order to compare the structure of Algorithm 19 to some of the most popular decomposition methods, we point out that the iterative steps can be reformulated for $n \geq 0$ as (here, $x^{-1} := x^0$)

$$\begin{aligned} y_i^{n+1} &:= \operatorname{prox}_{\sigma g_i^*}(y_i^n + \sigma(2x^n - x^{n-1})), i = 1, ..., k \\ x^{n+1} &:= x^n - \frac{\tau}{k}\sum_{i=1}^k y_i^{n+1} \end{aligned}$$

(i) The (parallel) Douglas-Rachford splitting algorithm, as described in [5, Proposition 25.7], applied to problem (36) gives rise for $n \geq 0$ to the following iterative steps

$$\begin{aligned} p^n &=: \frac{1}{k}\sum_{i=1}^k u_i^n \\ v_i^n &=: \operatorname{prox}_{\gamma g_i}(u_i^n), i = 1, ..., k \\ q^n &=: \frac{1}{k}\sum_{i=1}^k v_i^n \\ u_i^{n+1} &=: u_i^n + \lambda_n(2q^n - p^n - v_i^n), i = 1, ..., k \end{aligned}$$

Taking in the primal-dual algorithm $\sigma = \tau = 1$ and in the Douglas-Rachford splitting algorithm $\gamma = 1$ and $\lambda_n = 1$ for any $n \geq 0$, it turns out that the two iterative schemes are equivalent in the following sense. If the sequence $(x^n, y_1^n, ..., y_k^n)_{n\geq 0}$ is generated by the primal-dual algorithm, then the sequence $(u_1^n, ..., u_k^n, v_1^n, ..., v_k^n)_{n\geq 0}$, where $u_i^n := y_i^n + 2x^n - x^{n-1}$ and $v_i^n := u_i^n - y_i^{n+1}$, for $i = 1, ..., k$, is the one generated by the Douglas-Rachford splitting algorithm. Conversely, if $(u_1^n, ..., u_k^n, v_1^n, ..., v_k^n)_{n\geq 0}$ is the sequence generated by the Douglas-Rachford splitting algorithm, then $(x^n, y_1^n, ..., y_k^n)_{n\geq 0}$, where $y_i^n := u_i^{n-1} - v_i^{n-1}, i = 1, ..., k$, and $x^n := u_i^n - y_i^n + \frac{1}{k}\sum_{i=1}^k y_i^n$, is the one generated by the primal-dual algorithm.

(ii) The alternating direction method of multipliers (ADMM) applied to the dual optimization problem (37) gives rise (see [9, Subsection 7.3.2]) for $n \geq 0$ to the following iterative steps

$$\begin{aligned} y_i^{n+1} &=: \operatorname{prox}_{\rho^{-1} g_i^*}(y_i^n - u^n - \overline{x}^n), i = 1, ..., k \\ \overline{x}^{n+1} &=: \frac{1}{k}\sum_{i=1}^k y_i^{n+1} \\ u^{n+1} &=: u^n + \overline{x}^{n+1} \end{aligned}$$

which can be equivalently written as

$$\begin{aligned} y_i^{n+1} &=: \operatorname{prox}_{\rho^{-1} g_i^*}(y_i^n - (2u^n - u^{n-1})), i = 1, ..., k \\ u^{n+1} &=: u^n + \frac{1}{k}\sum_{i=1}^k y_i^{n+1}. \end{aligned}$$

By defining $x^n := -u^n$ for all $n \geq 0$, this is further equivalent to

$$\begin{aligned} y_i^{n+1} &=: \operatorname{prox}_{\rho^{-1} g_i^*}(y_i^n + (2x^n - x^{n-1})), i = 1, ..., k \\ x^{n+1} &=: x^n - \frac{1}{k}\sum_{i=1}^k y_i^{n+1} \end{aligned}$$

Thus, the primal-dual method applied to (36) for $\sigma = \tau = 1$ is equivalent to ADMM with penalty parameter $\rho = 1$ applied to the dual (37).

(iii) Spingarn's method, as described in [27, Section 5], applied to problem (36) gives rise for $n \geq 0$ to the following iterative steps (here, $\sum_{i=1}^k y_i^0 = 0$)

$$\begin{aligned} \overline{y}_i^n &=: \operatorname{prox}_{g_i^*}(x^n + y_i^n), i = 1, ..., k \\ \overline{x}_i^n &=: x^n + y_i^n - \overline{y}_i^n, i = 1, ..., k \\ x^{n+1} &=: \frac{1}{k}\sum_{i=1}^k \overline{x}_i^n \\ y_i^{n+1} &=: \overline{y}_i^n - \frac{1}{k}\sum_{i=1}^k \overline{y}_i^n, i = 1, ..., k \end{aligned}$$

which can be equivalently written as

$$\begin{aligned} \overline{y}_i^n &=: \operatorname{prox}_{g_i^*}(x^n + y_i^n), i = 1, ..., k \\ x^{n+1} &=: x^n - \frac{1}{k}\sum_{i=1}^k \overline{y}_i^n \\ y_i^{n+1} &=: \overline{y}_i^n - \frac{1}{k}\sum_{i=1}^k \overline{y}_i^n, i = 1, ..., k \end{aligned}$$

and further as

$$\begin{aligned} \overline{y}_i^{n+1} &=: \operatorname{prox}_{g_i^*}(\overline{y}_i^n + 2x^{n+1} - x^n), i = 1, ..., k. \\ x^{n+2} &=: x^{n+1} - \frac{1}{k}\sum_{i=1}^k \overline{y}_i^{n+1} \end{aligned}$$

One can recognize in the reformulation of Spingarn's method the same structure as for the primal-dual method in the case when $\sigma = 1$ and $\tau = 1$.

5 Numerical experiments

In this section we present numerical experiments involving the primal-dual algorithm and some of its variants when solving some nondifferentiable convex optimization problems originating in image processing and in location theory.

5.1 Image deblurring and denoising

The first numerical experiment concerns the solving of an ill-conditioned linear inverse problem arising in image deblurring. To this end, we consider images of size $M \times N$ as vectors $x \in \mathbb{R}^m$ for $m = MN$, where each pixel denoted by $x_{i,j}$, $1 \leq i \leq M$, $1 \leq j \leq N$, ranges in the closed interval from 0 (pure black) to 1 (pure white). For a given matrix $A \in \mathbb{R}^{m \times m}$ describing a blur operator and a given vector $b \in \mathbb{R}^m$ representing the blurred and noisy image the task that we considered was to estimate the unknown original image $x^* \in \mathbb{R}^m$ solving the linear system $Ax = b$.

To this aim we solved the following regularized convex nondifferentiable problem

$$\inf_{x \in [0,1]^m} \left\{ \|Ax - b\|^2 + \lambda_1 TV(x) + \lambda_2 \|x\|_1 \right\}, \tag{39}$$

where $TV : \mathbb{R}^m \to \mathbb{R}$ is a discrete total variation functional, $\lambda_1, \lambda_2 > 0$ are regularization parameters and the regularization is done by a combination of two functionals with different properties.

Two popular choices for the discrete total variation functional are the anisotropic total variation $TV_{\text{aniso}} : \mathbb{R}^m \to \mathbb{R}$

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

and the isotropic total variation $TV_{\text{iso}} : \mathbb{R}^m \to \mathbb{R}$

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

We denote $\mathcal{Y} = \mathbb{R}^m \times \mathbb{R}^m$ and define the linear operator $L : \mathbb{R}^m \to \mathcal{Y}$, $x_{i,j} \mapsto (L_1 x_{i,j}, L_2 x_{i,j})$, 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 in the horizontal and vertical directions. One can easily check that $\|L\|^2 \leq 8$ and that its adjoint $L^* : \mathcal{Y} \to \mathbb{R}^m$ is as easy to implement as the operator itself (cf. [11]).

We concretely looked at the 256×256 cameraman test image, which is part of the image processing toolbox in Matlab. We scaled the pixels to the interval $[0, 1]$ and vectorized the image, obtaining a vector of dimension $m = 256 \times 256 = 65536$. Further, by making use of the Matlab functions `imfilter` and `fspecial`, we blurred the image as follows:

```
H=fspecial('gaussian',9,4);     % gaussian blur of size 9 times 9
                                % and standard deviation 4
B=imfilter(X,H,'conv','symmetric');  % B=observed blurred image
                                     % X=original image
```

In row 1 the function `fspecial` returns a rotationally symmetric Gaussian lowpass filter of size 9×9 with standard deviation 4. The entries of H are nonnegative and their sum adds up to 1. In row 3 the function `imfilter` convolves the filter H with the image X and outputs the blurred image B. The boundary option "symmetric" corresponds to reflexive boundary conditions.

Thanks to the rotationally symmetric filter H, the linear operator $A \in \mathbb{R}^{m \times m}$ given by the Matlab function `imfilter` is symmetric, too. By making use of the real spectral decomposition of A it shows that $\|A\|^2 = 1$. After adding a zero-mean white Gaussian noise with standard deviation 10^{-3}, we obtained the blurred and noisy image $b \in \mathbb{R}^m$ which is shown in Figure 1. We solved (39) by applying Algorithm 17 for both instances

original | blurred and noisy

Figure 1: The 256×256 cameraman test image

of the discrete total variation functional. Thus, when considering the anoisotropic total variation, the problem (39) can be formulated as

$$\inf_{x \in \mathbb{R}^m} \left\{ g_1(Ax) + g_2(Lx) + g_3(x) + g_4(x) \right\}, \tag{40}$$

where $g_1 : \mathbb{R}^m \to \mathbb{R}, g_1(x) = \|x - b\|^2$, $g_2 : \mathcal{Y} \to \mathbb{R}$, $g_2(u,v) = \lambda_1 \|(u,v)\|_1$, $g_3 : \mathbb{R}^m \to \mathbb{R}$, $g_3(x) = \lambda_2 \|x\|_1$ and $g_4 : \mathbb{R}^m \to \overline{\mathbb{R}}$ is the indicator function of $[0,1]^m$.

When applying standard splitting algorithms in order to solve problem (40), one would need determing the resolvents of $A^* \circ \partial g_1 \circ A$ and $L^* \circ \partial g_2 \circ L$, for which in general no exact formulae are available. On the other hand, the proposed primal-dual method asks only for the proximal points of the conjugates of the particular functions, which, as it is shown as follows, can be easily calculated.

For every $p \in \mathbb{R}^m$ we have $g_1^*(p) = \frac{1}{4}\|p\|^2 + p^T b$ and $g_3^*(p) = \delta_{[-\lambda_2,\lambda_2]^m}(p)$, while, for every $(p,q) \in \mathcal{Y}$ it holds $g_2^*(p,q) = \delta_{[-\lambda_1,\lambda_1]^m \times [-\lambda_1,\lambda_1]^m}(p,q)$. Thus, for $\sigma > 0$ and $p, q \in \mathbb{R}^m$ we have

$$\begin{aligned} \operatorname{prox}_{\sigma g_1^*}(p) &= \frac{2(p - b\sigma)}{2+\sigma}, \\ \operatorname{prox}_{\sigma g_2^*}(p,q) &= P_{[-\lambda_1,\lambda_1]^m \times [-\lambda_1,\lambda_1]^m}(p,q), \\ \operatorname{prox}_{\sigma g_3^*}(p) &= P_{[-\lambda_2,\lambda_2]^m}(p) \end{aligned}$$

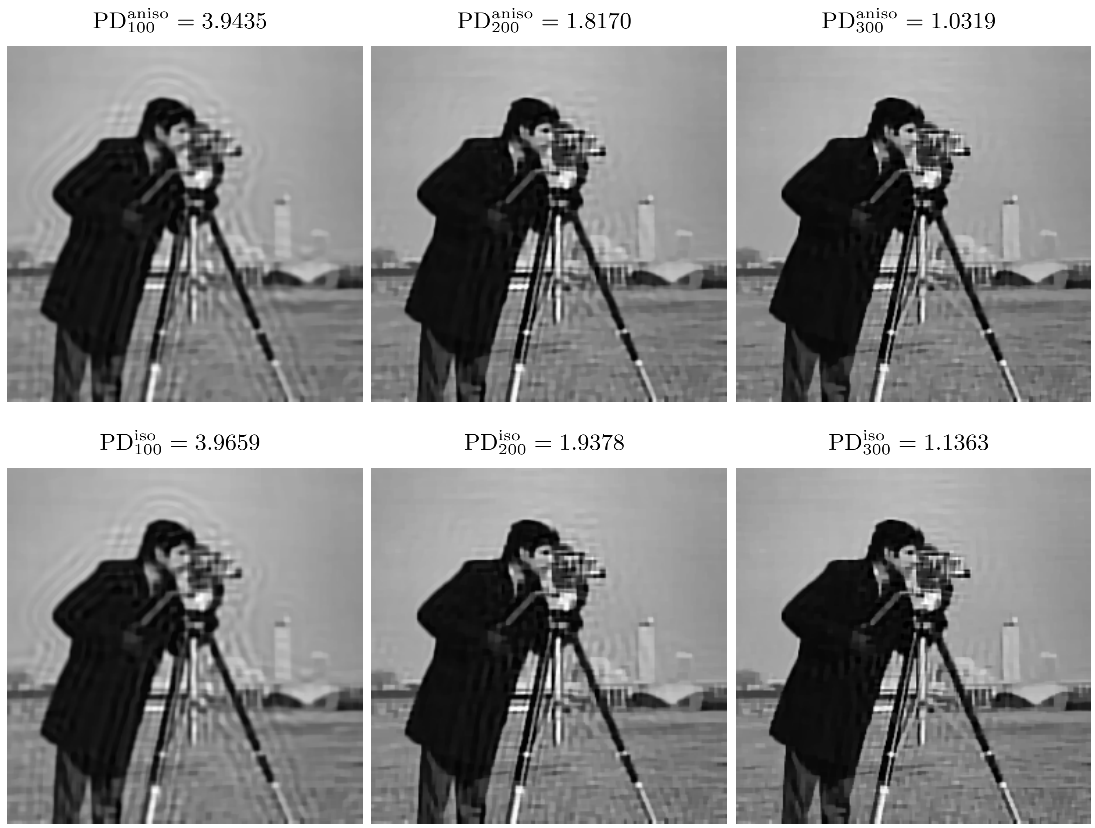

Figure 2: Iterations 100, 200 and 300 when solving the problems (40) and (41) via Algorithm 17

and

$$\mathrm{prox}_{\sigma g_4^*}(p) = p - \sigma\, \mathrm{prox}_{\sigma \delta_{[0,1]^m}}\left(\tfrac{1}{\sigma}p\right) = p - \sigma P_{[0,1]^m}\left(\tfrac{1}{\sigma}p\right).$$

On the other hand, when considering the isotropic total variation, the problem (39) can be formulated as

$$\inf_{x\in\mathbb{R}^m} \left\{ g_1(Ax) + \tilde{g}_2(Lx) + g_3(x) + g_4(x) \right\}, \tag{41}$$

where the functions g_1, g_3 and g_4 are taken as above and $\tilde{g}_2 : \mathcal{Y} \to \mathbb{R}$ is defined as $\tilde{g}_2(u,v) = \lambda_1 \|(u,v)\|_\times$, where $\|(\cdot,\cdot)\|_\times : \mathcal{Y} \to \mathbb{R}$, $\|(u,v)\|_\times = \sum_{i=1}^M \sum_{j=1}^N \sqrt{u_{i,j}^2 + v_{i,j}^2}$, is a norm on the Hilbert space $\mathcal{Y}$. Thus for every $(p,q) \in \mathcal{Y}$ it holds $\tilde{g}_2^*(p,q) = \delta_S(p,q)$ and

$$\mathrm{prox}_{\sigma \tilde{g}_2^*}(p,q) = P_S\left(p,q\right),$$

where (cf. [8])

$$S = \left\{ (p,q) \in \mathcal{Y} : \max_{\substack{1\leq i\leq M\\ 1\leq j\leq N}} \sqrt{p_{i,j}^2 + q_{i,j}^2} \leq \lambda_1 \right\}$$

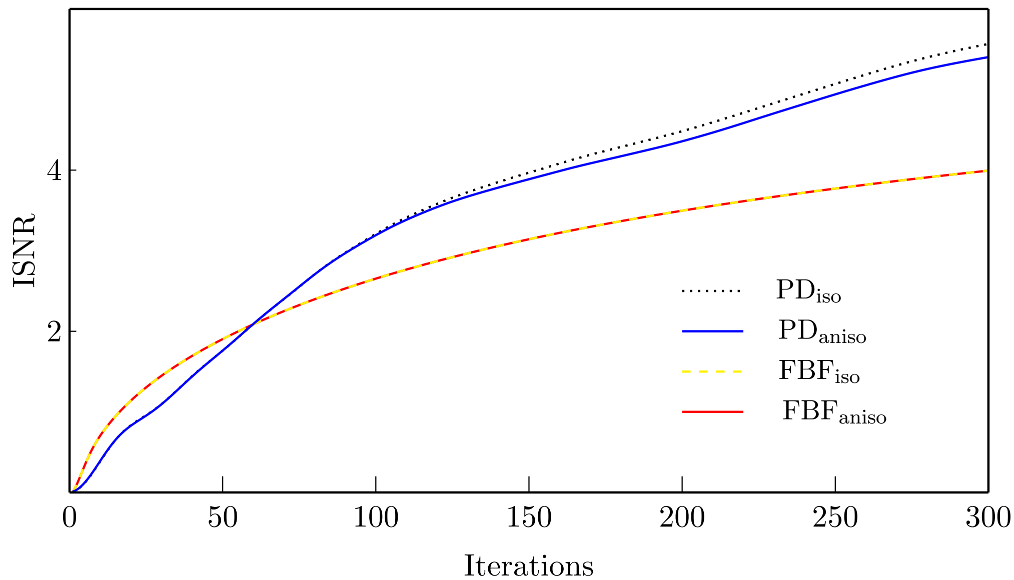

Figure 3: Improvement in signal-to-noise ratio (ISNR)

and the projection operator $P_S : \mathcal{Y} \to S$ is defined via

$$(p_{i,j}, q_{i,j}) \mapsto \lambda_1 \frac{(p_{i,j}, q_{i,j})}{\max\left\{\lambda_1, \sqrt{p_{i,j}^2 + q_{i,j}^2}\right\}},\ 1 \leq i \leq M,\ 1 \leq j \leq N.$$

We considered as regularization parameter $\lambda_1 = \lambda_2 = 2 \times 10^{-5}$, $\sigma = 64 \times 10^{-5}$ and $\tau = 140$ when solving (40) and $\sigma = 9 \times 10^{-4}$ and $\tau = 100$ when solving (41). The top line of Figure 2 shows the iterations 100, 200 and 300 when solving (40) via Algorithm 17, while the bottom line of it shows the iterations 100, 200 and 300 when solving (41) via the same algorithm, for each of them the value of the objective function at the respective iterate being provided.

We also made some comparisons from the point of view of the quality of the recovered images when solving (40) and (41) with Algorithm 17 and with the iterative method from [10,13] relying on Tseng's Forward-Backward-Forward algorithm. We considered the latter not only because of its primal-dual nature, but also since it assumes only simple forward evaluations of the linear continuous operators (and of their adjoints) involved in the formulation of the optimization problems. For the iterative scheme from [10, 13] we have chosen $\varepsilon = 0.012$ and $\gamma = 0.3124$. The comparisons concerning the quality of the recovered images were made via the improvement in signal-to-noise ratio (ISNR), which is defined as

$$\mathrm{ISNR}(n) = 10 \log_{10} \left(\frac{\|x - b\|^2}{\|x - x^n\|^2} \right),$$

where x, b and x^n denote the original, observed and estimated image at iteration n, respectively. Figure 3 and Figure 4 show the evolution of the ISNR values and the decrease in the objective function values, respectively, when solving (40) and (41) with the Algorithm 17 (PD) and with the algorithm from [10,13] (FBF), respectively. Algorithm 17 furnishes

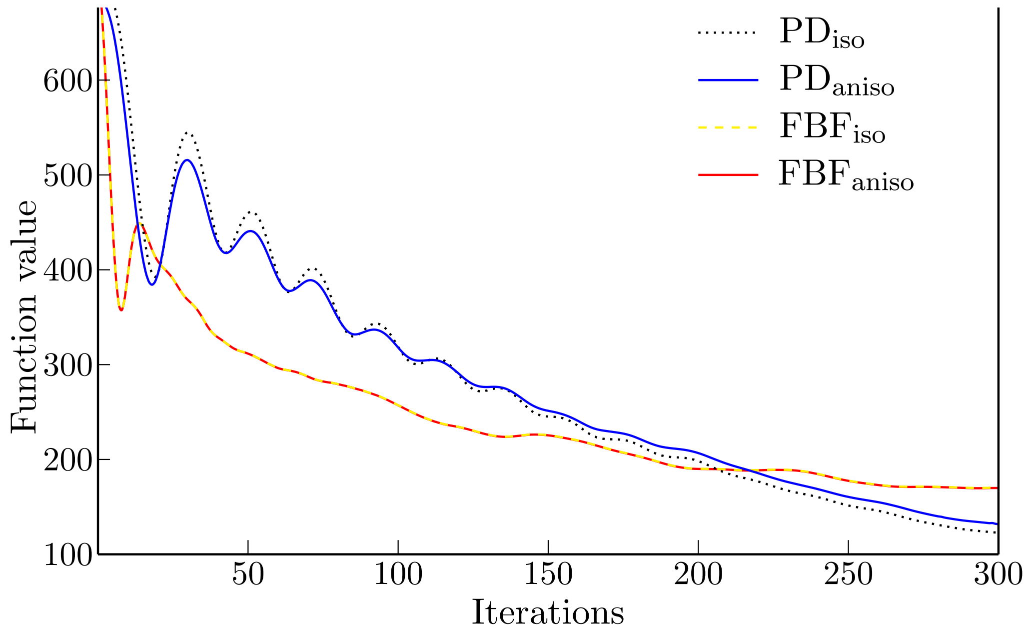

Figure 4: The decrease in the objective function values

after 300 iterations from the point of view of the quality of the recovered images the best results, whereby here the isotropic total variation functional is the optimal choice. More than that, as expected, due to the supplementary forward step, the algorithm in [10,13] needs at each iterate more CPU time than Algorithm 17.

5.2 The Fermat-Weber problem

The second application of the primal-dual algorithm presented in this paper is with respect to the solving of the Fermat-Weber problem, which concerns the finding of a new facility in order to minimize the sum of weighted distances to a set of fixed points. We considered the nondifferentiable convex optimization problem

$$(P_{FW}) \qquad \inf_{x\in\mathbb{R}^m}\left\{\sum_{i=1}^{k}\lambda_i\|x-c_i\|\right\},$$

where $c_i \in \mathbb{R}^m$ are given points and $\lambda_i > 0$ are given weights for $i = 1, ..., k$. We solved the optimization problem (P_{FW}) by using Algorithm 19 for $g_i : \mathbb{R}^m \to \mathbb{R}, g_i(x) = \lambda_i\|x-c_i\|, i = 1, ..., k$. With this respect we used that for $i = 1, ..., k$ it holds

$$g_i^*(y) = \begin{cases} \langle y, c_i\rangle, & \text{if } \|y\| \leq \lambda_i, \\ +\infty, & \text{otherwise,} \end{cases} \quad \forall y \in \mathbb{R}^m$$

and, from here, when $\sigma > 0$, that

$$\operatorname{prox}_{\sigma g_i^*}(z) = \begin{cases} z-\sigma c_i, & \text{if } \|z-\sigma c_i\| \leq \lambda_i, \\ \lambda_i\frac{z-\sigma c_i}{\|z-\sigma c_i\|}, & \text{otherwise} \end{cases} \quad \forall z \in \mathbb{R}^m.$$

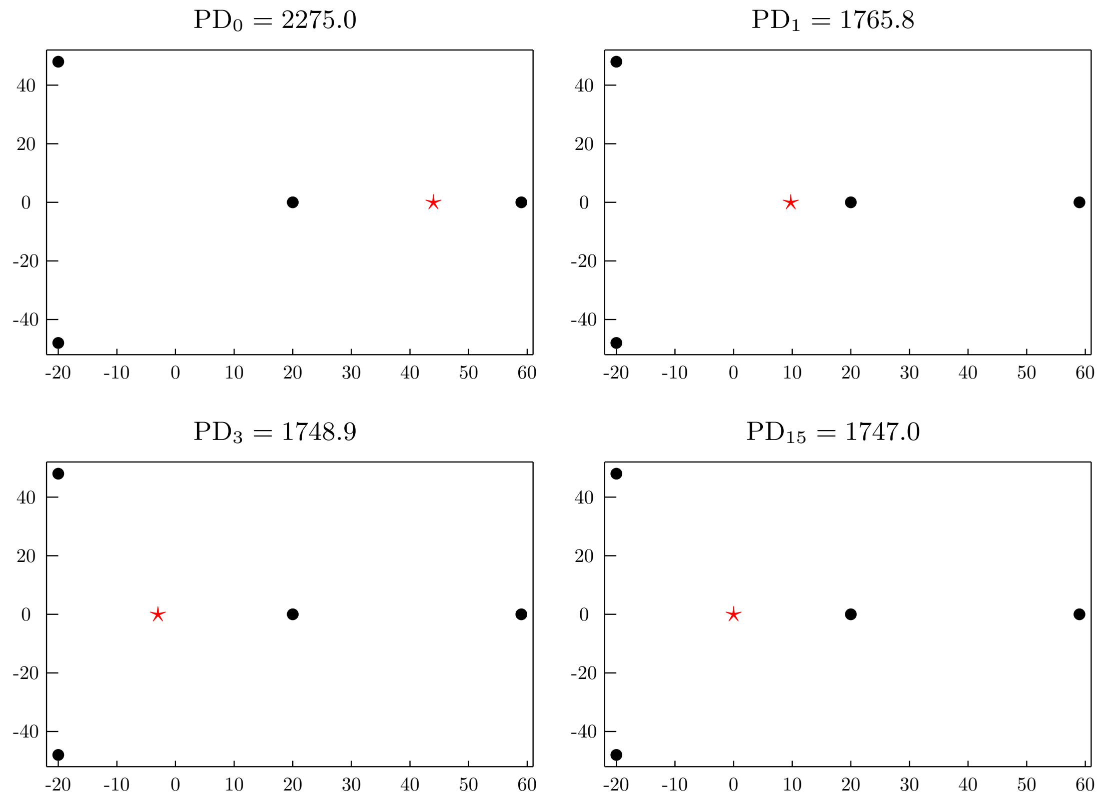

Figure 5: The progression of iterations of Algorithm 19 when solving the Fermat-Weber problem for points and weights given by (42)

We investigated the functionality of the algorithm on two sets of points and weights, often considered in the literature when analyzing the performances of iterative schemes for the Fermat-Weber problem. In the light of the theoretical considerations made at the end of the previous section, the convergence behaviour of classical splitting algorithms will not differ from the one of the primal-dual method. Thus, we will compare the performances of the primal-dual algorithm with the ones of a smoothing method recently proposed in the literature and tested in the context of the the Fermat-Weber problem. In the last years one can notice a dramatically increase of the number of publications dealing with smoothing approaches for solving nonsmooth convex optimization problems. They become popular since they provide good convergence rates, especially for the sequence of objective values, motivated by the fact that they rely on fast gradient methods. By considering the following two examples we would like to emphasize the fact that, in counterpart to the splitting algorithms, they seem not to be appropriate when applied to location problems.

In a first instance we considered for $k = 4$ the points in the plane and the weights

$$c_1 = (59, 0), c_2 = (20, 0), c_3 = (-20, 48), c_4 = (-20, -48) \text{ and } \lambda_1 = \lambda_2 = 5, \lambda_3 = \lambda_4 = 13, \tag{42}$$

respectively. The optimal location point is $\widehat{x} = (0, 0)$, however, the classical Weiszfeld algorithm (see [22, 30]) with starting point $x^0 = (44, 0)$ breaks down, due to the fact that

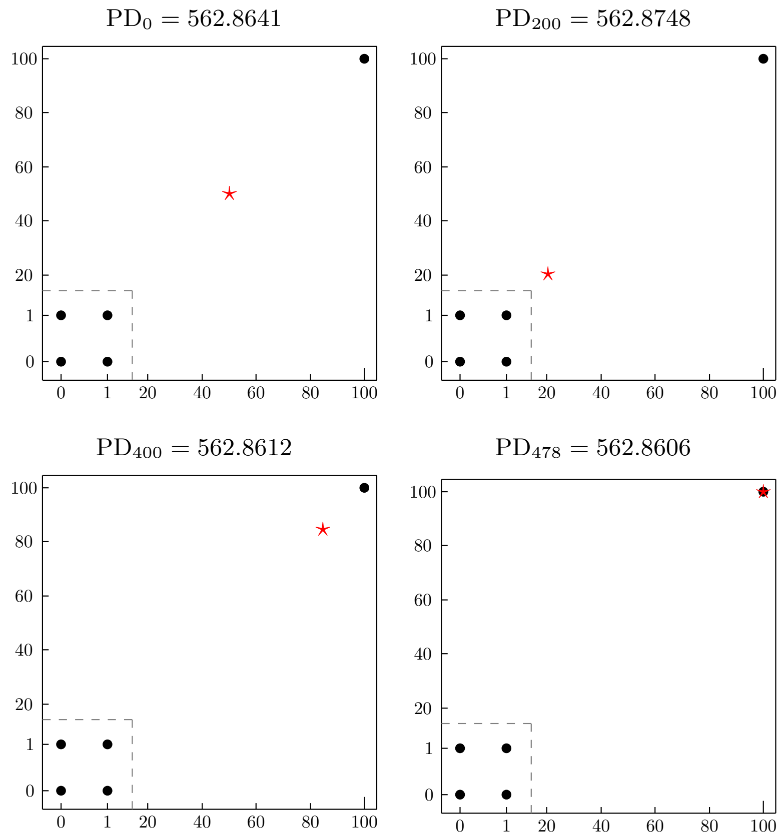

Figure 6: The progression of iterations of Algorithm 19 when solving the Fermat-Weber problem for points and weights given by (43)

the sequence of generated iterates attains the point $(20, 0)$. On the other hand, Algorithm 19 with $\sigma = 0.13$, $\tau = 7.6923$ and $y_k^0 = (0, 0), k = 1, ..., 4$, achieved a point which is optimal up to three decimal points after 15 iterations. Figure 5 shows the progression of the iterations, while PD_n provides the value of the objective function at iteration n.

Recently, an approach for solving the Fermat-Weber problem was proposed by Goldfarb and Ma in [21], which assumes the approximation of each of the functions in the objective by a convex and differentiable function with Lipschitz-continuous gradient. The optimization problem which this smoothing method yields is solved in [21] by the classical gradient method (Grad) and by a variant of Nesterov's accelerated gradient method (Nest) (see [23]) and by a fast multiple-splitting algorithm (FaMSA-s) introduced in this paper. We applied the smoothing approach in connection with these algorithms to the example considered in (42) with smoothness parameter ρ equal to 10^{-3} (chosen also in [21]) and step sizes $\tau = 0.1$, $\tau = 0.01$ and $\tau = 0.001$. We stopped the three algorithms when achieving an iterate x^n such that $\|x^n - \widehat{x}\| \leq 10^{-3}$ and obtained in all cases the lowest number of iterations for $\tau = 0.1$. A point which is optimal up to three decimal points was obtained for Nest after 308 iterations, for Grad after 175 iterations and for FaMSA-s after 54 iterations, thus none of these iterative schemes attained the performance of Algorithm 19.

For the second example of the Fermat-Weber problem we considered in case $k = 5$ the points in the plane and the weights (see [15])

$$\begin{aligned} c_1 = (0,0), c_2 = (1,0), c_3 = (0,1),\ c_4 = (1,1), c_5 = (100,100) \\ \text{and } \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1,\ \lambda_5 = 4, \end{aligned} \tag{43}$$

respectively. The optimal location point is $\widehat{x} = (100, 100)$ and, by choosing the relative center of gravity $x^0 = (50.25, 50.25)$ as starting point, we found out that not only the classical Weiszfeld algorithm, but also the approach from [21] described above in connection to each of the methods Grad, Nest and FaMSA-s did not achieve a point which is optimal up to three decimal points after millions of iterations. On the other hand, Algorithm 19 with $\sigma = 0.0001$, $\tau = 9999$ and $y_k^0 = (0,0), k = 1, ..., 5$, achieved a point which is optimal up to three decimal points after 478 iterations. This example is more than illustrative for the performance of the primal-dual Algorithm 19 in comparison to some classical and recent algorithms designed for the Fermat-Weber problem. Figure 6 shows the progression of the iterations, while PD_n provides the value of the objective function at iteration n.

Acknowledgements. The authors are thankful to Christopher Hendrich for the implementation of the numerical schemes to which comparisons of the primal-dual algorithm were made. The authors are also thankful to Heinz Bauschke and to anonymous reviewers for recommendations and remarks which improved the quality of the paper.

References

[1] G. Alduncin, *Composition duality principles for mixed variational inequalities*, Mathematical and Computer Modelling 41(6-7), 639–654, 2005

[2] H. Attouch, M. Théra, *A general duality principle for the sum of two operators*, Journal of Convex Analysis 3(1), 1-24, 1996

[3] K.J. Arrow, L. Hurwicz, H. Uzawa, *Studies in linear and nonlinear programming*, In: Cheney, H.B., Johnson, S.M., Karlin, S., Marschak, T., Solow, R.M. (eds.) "Stanford Mathematical Studies in the Social Sciences", Vol. II, Stanford University Press, Stanford, 1958

[4] H.H. Bauschke, R.I. Boţ, W.L. Hare, W.M. Moursi, *Attouch-Théra duality revisited: paramonotonicity and operator splitting*, Journal of Approximation Theory 164(8), 1065–1084, 2012

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

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

[7] 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

[8] R.I. Boţ, C. Hendrich, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, arXiv:1211.1706v1 [math.OC], 2012

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning 3(1), 1–122, 2010

[10] 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

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

[12] 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

[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] J. Douglas, H.H. Rachford, *On the numerical solution of the heat conduction problem in* 2 *and* 3 *space variables*, Transactions of the American Mathematical Society 82, 421–439, 1956

[15] Z. Drezner, *A note on accelerating the Weiszfeld procedure*, Location Science 3(4), 275–279, 1995

[16] J. Eckstein, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, Rutcor Research Report 32-2012, 2012

[17] J. Eckstein, D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, 293–318, 1992

[18] J. Eckstein, M.C. Ferris, *Smooth methods of multipliers for complementarity problems*, Mathematical Programming 86(1), 65–90, 1999

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

[20] E. Esser, X. Zhang, T.F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences 3(4), 1015-1046, 2010

[21] D. Goldfarb, S. Ma, *Fast multiple-splitting algorithms for convex optimization*, SIAM Journal on Optimization 22(2), 533–556, 2012

[22] R.F. Love, J.G. Morris, G.O. Wesolowsky, *Facilities Location Models and Methods*, North-Holland Publishing Company, Amsterdam, 1988

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

[24] T. Pennanen, *Dualization of generalized equations of maximal monotone type*, SIAM Journal on Optimization 10(3), 809–835, 2000

[25] R.T. Rockafellar, *On the maximal monotonicity of subdiferential mappings*, Pacific Journal of Mathematics 33(1), 209–216, 1970

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

[27] J.E. Spingarn, *Partial inverse of a monotone operator*, Applied Mathematics and Optimization 10, 247–265, 1983

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

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

[30] E. Weiszfeld, *Sur le point pour lequel la somme des distances de n points donnes est minimum*, Tohoku Mathematical Journal 43, 597–609, 1937

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