

# Envelope Functions: Unifications and Further Properties

Pontus Giselsson<sup>1</sup>  · Mattias Fält<sup>1</sup>

Received: 3 February 2017 / Accepted: 4 June 2018 / Published online: 12 June 2018  
© The Author(s) 2018

**Abstract** Forward–backward and Douglas–Rachford splitting are methods for structured nonsmooth optimization. With the aim to use smooth optimization techniques for nonsmooth problems, the forward–backward and Douglas–Rachford envelopes were recently proposed. Under specific problem assumptions, these envelope functions have favorable smoothness and convexity properties and their stationary points coincide with the fixed-points of the underlying algorithm operators. This allows for solving such nonsmooth optimization problems by minimizing the corresponding smooth convex envelope function. In this paper, we present a general envelope function that unifies and generalizes existing ones. We provide properties of the general envelope function that sharpen corresponding known results for the special cases. We also present a new interpretation of the underlying methods as being majorization–minimization algorithms applied to their respective envelope functions.

**Keywords** First-order methods · Envelope functions · Nonsmooth optimization · Smooth reformulations · Large-scale optimization

**Mathematics Subject Classification** 90C30 · 47J25

---

Communicated by Nicolas Hadjisavvas.

---

✉ Pontus Giselsson  
pontusg@control.lth.se

Mattias Fält  
mattiasf@control.lth.se

<sup>1</sup> Department of Automatic Control, Lund University, Box 118, SE-221 00 Lund, Sweden

## 1 Introduction

Many convex optimization problems can be reformulated into a problem of finding a fixed-point of a nonexpansive operator. This is the basis for many first-order optimization algorithms such as; forward–backward splitting [1], Douglas–Rachford splitting [2, 3], the alternating direction method of multipliers (ADMM) [4–6] and its linearized versions [7], the three operator splitting method [8], and (generalized) alternating projections [9–14].

In these methods, a fixed-point is found by performing an averaged iteration of the nonexpansive mapping. This scheme guarantees global convergence, but the rate of convergence can be slow. A well studied approach for improving practical convergence—that has proven very successful in practice—is preconditioning of the problem data; see, e.g., [15–21] for a limited selection of such methods. The underlying idea is to incorporate static second-order information in the respective algorithms.

The performance of the forward–backward and the Douglas–Rachford methods can be further improved by exploiting the properties of the recently proposed forward–backward envelope [22, 23] and Douglas–Rachford envelope [24]. As shown in [22–24], the stationary points of these envelope functions agree with the fixed-points of the corresponding algorithm operator. Under certain assumptions, they have favorable properties such as convexity and Lipschitz continuity of the gradient. These properties enable for nonsmooth problems to be solved by finding a stationary point of a smooth and convex envelope function. In [22, 23], truncated Newton methods and quasi-Newton methods are applied to the forward–backward envelope function to improve local convergence. During the submission procedure of this paper, these works have been extended to the nonconvex setting in [25, 26] for both forward–backward splitting and Douglas–Rachford splitting.

A unifying property of forward–backward and Douglas–Rachford splitting (for convex optimization) is that they are averaged iterations of a nonexpansive mapping. This mapping is composed of two nonexpansive mappings that are gradients of functions. Based on this observation, we present a general envelope function that has the forward–backward envelope and the Douglas–Rachford envelope as special cases. Other special cases include the Moreau envelope and the ADMM envelope [27], since they are special cases of the forward–backward and Douglas–Rachford envelopes respectively. We also explicitly characterize the relationship between the ADMM and Douglas–Rachford envelopes as being essentially the negatives of each other.

The analyses of the envelope functions in [22–24] require, translated to our setting, that one of the functions that define one of the nonexpansive operators in the composition, is twice continuously differentiable. In this paper, we analyze the proposed general envelope function in the more restrictive setting of the twice continuously function being quadratic, or equivalently its gradient being affine. We show that if the Hessian matrix of this function is nonsingular the stationary points of the envelope coincide with the fixed-points of the nonexpansive operator. We provide sharp quadratic upper and lower bounds to the envelope function that improve corresponding results for the known special cases in the literature. One implication of these bounds is that the gradient of the envelope function is Lipschitz continuous with constant two. If, in addition, the before mentioned Hessian matrix is positive semidefinite the envelope

function is convex, implying that a fixed-point to the nonexpansive operator can be found by minimizing a smooth and convex envelope function.

We also provide an interpretation of the basic averaged fixed-point iteration as a majorization–minimization step on the envelope function. We show that the majorizing function is a quadratic upper bound, which is slightly more conservative than the provided sharp quadratic upper bound. We also note that using the sharp quadratic upper bound as majorizing function would result in computationally more expensive algorithm iterations.

Our contributions are as follows; (i) we propose a general envelope function that has several known envelope functions as special cases, (ii) we provide properties of the general envelope that sharpen (sometimes considerably) and generalize corresponding known results for the special cases, (iii) we provide an interpretation of the basic averaged iteration as a suboptimal majorization–minimization step on the envelope (iv) we provide new insights on the relation between the Douglas–Rachford envelope and the ADMM envelope.

## 2 Preliminaries

### 2.1 Notation

We denote by  $\mathbb{R}$  the set of real numbers,  $\mathbb{R}^n$  the set of real  $n$ -dimensional vectors, and  $\mathbb{R}^{m \times n}$  the set of real  $m \times n$ -matrices. Further  $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$  denotes the extended real line. We denote inner-products on  $\mathbb{R}^n$  by  $\langle \cdot, \cdot \rangle$  and their induced norms by  $\| \cdot \|$ . We define the scaled norm  $\|x\|_P := \sqrt{\langle Px, x \rangle}$ , where  $P$  is a positive definite operator (defined in Definition 2.2). We will use the same notation for scaled semi-norms, i.e.,  $\|x\|_P := \sqrt{\langle Px, x \rangle}$ , where  $P$  is a positive semidefinite operator (defined in Definition 2.1). The identity operator is denoted by  $\text{Id}$ . The conjugate function is denoted and defined by  $f^*(y) \triangleq \sup_x \{\langle y, x \rangle - f(x)\}$ . The adjoint operator to a linear operator  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined as the unique operator  $L^* : \mathbb{R}^m \rightarrow \mathbb{R}^n$  that satisfies  $\langle Lx, y \rangle = \langle x, L^*y \rangle$ . The linear operator  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is self-adjoint if  $L = L^*$ . The notation  $\text{argmin}_x f(x)$  refers to any element that minimizes  $f$ . Finally,  $\iota_C$  denotes the indicator function for the set  $C$  that satisfies  $\iota_C(x) = 0$  if  $x \in C$  and  $\iota_C(x) = \infty$  if  $x \notin C$ .

### 2.2 Background

In this section, we introduce some standard definitions that can be found, e.g., in [28, 29].

#### 2.2.1 Operator Properties

**Definition 2.1** (Positive semidefinite) A linear operator  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is *positive semidefinite*, if it is self-adjoint and all eigenvalues  $\lambda_i(L) \geq 0$ .

**Remark 2.1** An equivalent characterization of a positive semidefinite operator is that  $\langle Lx, x \rangle \geq 0$  for all  $x \in \mathbb{R}^n$ .

**Definition 2.2** (*Positive definite*) A linear operator  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is *positive definite*, if it is self-adjoint and if all eigenvalues  $\lambda_i(L) \geq m$  with  $m > 0$ .

**Remark 2.2** An equivalent characterization of a positive definite operator  $L$  is that  $\langle Lx, x \rangle \geq m\|x\|^2$  for some  $m > 0$  and all  $x \in \mathbb{R}^n$ .

**Definition 2.3** (*Lipschitz continuous*) A mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\delta$ -Lipschitz continuous with  $\delta \geq 0$  if

$$\|Tx - Ty\| \leq \delta\|x - y\|$$

holds for all  $x, y \in \mathbb{R}^n$ . If  $\delta = 1$ , then  $T$  is *nonexpansive* and if  $\delta \in [0, 1[$ , then  $T$  is  $\delta$ -contractive.

**Definition 2.4** (*Averaged*) A mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\alpha$ -averaged if there exists a nonexpansive mapping  $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and an  $\alpha \in ]0, 1]$  such that  $T = (1 - \alpha)\text{Id} + \alpha S$ .

**Definition 2.5** (*Negatively averaged*) A mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\beta$ -negatively averaged with  $\beta \in ]0, 1]$  if  $-T$  is  $\beta$ -averaged.

**Remark 2.3** For notational convenience, we have included  $\alpha = 1$  and  $\beta = 1$  in the definitions of (negative) averagedness, which both are equivalent to nonexpansiveness. For values of  $\alpha \in ]0, 1[$  and  $\beta \in ]0, 1[$  averagedness is a stronger property than nonexpansiveness. For more on negatively averaged operators, see [21] where they were introduced.

If a gradient operator  $\nabla f$  is  $\alpha$ -averaged and  $\beta$ -negatively averaged, then it must hold that  $\alpha + \beta \geq 1$ . This follows immediately from Lemma 3.1.

**Definition 2.6** (*Cocoerciveness*) A mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\delta$ -cocoercive with  $\delta > 0$  if  $\delta T$  is  $\frac{1}{2}$ -averaged.

**Remark 2.4** This definition implies that cocoercive mappings  $T$  can be expressed as

$$T = \frac{1}{2\delta}(\text{Id} + S), \quad (1)$$

where  $S$  is a nonexpansive operator. Therefore, 1-cocoercivity is equivalent to  $\frac{1}{2}$ -averagedness (which is also called firm nonexpansiveness).

### 2.2.2 Function Properties

**Definition 2.7** (*Strongly convex*) Let  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be positive definite. A proper and closed function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is  $\sigma$ -strongly convex w.r.t.  $\|\cdot\|_P$  with  $\sigma > 0$  if  $f - \frac{\sigma}{2}\|\cdot\|_P^2$  is convex.

*Remark 2.5* If  $f$  is differentiable,  $\sigma$ -strong convexity w.r.t.  $\|\cdot\|_P$  can equivalently be defined as that

$$\frac{\sigma}{2} \|x - y\|_P^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \tag{2}$$

holds for all  $x, y \in \mathbb{R}^n$ . If  $P = \text{Id}$ , i.e., if the norm is the induced norm, we merely say that  $f$  is  $\sigma$ -strongly convex. If  $\sigma = 0$ , the function is convex.

There are many smoothness definitions for functions in the literature. We will use the following, which describes the existence of majorizing and minimizing quadratic functions.

**Definition 2.8** (*Smooth*) Let  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be positive semidefinite. A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is  $\beta$ -smooth w.r.t.  $\|\cdot\|_P$  with  $\beta \geq 0$  if it is differentiable and

$$-\frac{\beta}{2} \|x - y\|_P^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{\beta}{2} \|x - y\|_P^2 \tag{3}$$

holds for all  $x, y \in \mathbb{R}^n$ .

### 2.2.3 Connections

Our main result (see Theorem 3.1) is that the envelope function satisfies upper and lower bounds of the form

$$\frac{1}{2} \langle M(x - y), x - y \rangle \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{1}{2} \langle L(x - y), x - y \rangle \tag{4}$$

for all  $x, y \in \mathbb{R}^n$  and for different linear operators  $M, L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Depending on  $M$  and  $L$ , we get different properties of  $f$  and its gradient  $\nabla f$ . Some of these are stated below. The results follow immediately from Lemma D.2 in Appendix D and the definitions of smoothness and strong convexity in Definitions 2.7 and 2.8, respectively.

**Proposition 2.1** Assume that  $L = -M = \beta I$  with  $\beta \geq 0$  in (4). Then, (4) is equivalent to that  $\nabla f$  is  $\beta$ -Lipschitz continuous.

**Proposition 2.2** Assume that  $M = \sigma I$  and  $L = \beta I$  with  $0 \leq \sigma \leq \beta$  in (4). Then, (4) is equivalent to that  $\nabla f$  is  $\beta$ -Lipschitz continuous and  $f$  is  $\sigma$ -strongly convex.

**Proposition 2.3** Assume that  $L = -M$  and that  $L$  is positive definite. Then, (4) is equivalent to that  $f$  is 1-smooth w.r.t.  $\|\cdot\|_L$ .

**Proposition 2.4** Assume that  $M$  and  $L$  are positive definite. Then, (4) is equivalent to that  $f$  is 1-smooth w.r.t.  $\|\cdot\|_L$  and 1-strongly convex w.r.t.  $\|\cdot\|_M$ .

### 3 Envelope Function

In [22, 24], the forward–backward and Douglas–Rachford envelope functions are proposed. Under certain problem data assumptions, these envelope functions have favorable properties; they are convex, they have Lipschitz continuous gradients, and their minimizers are fixed-points of the nonexpansive operator  $S$  that defines the respective algorithms. In this section, we will present a general envelope function that has the forward–backward and Douglas–Rachford envelopes as special cases. We will also provide properties of the general envelope that are sharper than what is known for the special cases.

We assume that the nonexpansive operator  $S$  that defines the algorithm is a composition of  $S_1$  and  $S_2$ , i.e.,  $S = S_2S_1$ , where  $S_1$  and  $S_2$  satisfy the following basic assumptions (that sometimes will be sharpened or relaxed).

**Assumption 3.1** Suppose that:

- (i)  $S_1 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $S_2 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  are nonexpansive.
- (ii)  $S_1 = \nabla f_1$  and  $S_2 = \nabla f_2$  for some differentiable functions  $f_1 : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ .
- (iii)  $f_1 : \mathbb{R}^n \rightarrow \mathbb{R}$  is twice continuously differentiable.

These assumptions are met for our algorithms of interest, see Sect. 4 for details. In this general framework, we propose the following envelope function:

$$F(x) := \langle \nabla f_1(x), x \rangle - f_1(x) - f_2(\nabla f_1(x)), \quad (5)$$

which has gradient

$$\begin{aligned} \nabla F(x) &= \nabla^2 f_1(x)x + \nabla f_1(x) - \nabla f_1(x) - \nabla^2 f_1(x)\nabla f_2(\nabla f_1(x)) \\ &= \nabla^2 f_1(x)(x - \nabla f_2(\nabla f_1(x))) \\ &= \nabla^2 f_1(x)(x - S_2S_1x). \end{aligned} \quad (6)$$

If the Hessian  $\nabla^2 f_1(x)$  is nonsingular for all  $x$ , then the set of stationary points of the envelope coincides with the fixed-points of  $S_2S_1$ .

**Proposition 3.1** Suppose that Assumption 3.1 holds and that  $\nabla^2 f(x)$  is nonsingular for all  $x \in \mathbb{R}^n$ . Let

$$X^* := \{x \in \mathbb{R}^n : \nabla F(x) = 0\}, \quad \text{fix}(S_2S_1) = \{x \in \mathbb{R}^n : S_2S_1x = x\}.$$

Then,  $X^* = \text{fix}(S_2S_1)$ .

*Proof* The statement follows trivially from (6).  $\square$

In Sect. 4, we show that the forward–backward and Douglas–Rachford envelopes are special cases of (5). In this section, we will provide properties of the general envelope under the following restriction to Assumption 3.1.

**Assumption 3.2** Suppose that Assumption 3.1 holds and that, in addition,  $S_1 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is affine, i.e.,  $S_1x = Px + q$  and  $f_1(x) = \frac{1}{2}\langle Px, x \rangle + \langle q, x \rangle$ , where  $P \in \mathbb{R}^{n \times n}$  is a self-adjoint nonexpansive linear operator and  $q \in \mathbb{R}^n$ .

*Remark 3.1* That  $P$  a self-adjoint nonexpansive linear operator means that it is symmetric with eigenvalues in the interval  $[-1, 1]$ .

When  $S_1 = \nabla f_1 = P(\cdot) + q$  is affine, the first two terms in the envelope function definition in (5) satisfy

$$\langle \nabla f_1(x), x \rangle - f_1(x) = \langle Px + q, x \rangle - \left( \frac{1}{2}\langle Px, x \rangle + \langle q, x \rangle \right) = \frac{1}{2}\langle Px, x \rangle.$$

Therefore, the general envelope function in (5) reduces to

$$F(x) = \frac{1}{2}\langle Px, x \rangle - f_2(\nabla f_1(x)) \tag{7}$$

and its gradient (6) becomes

$$\nabla F(x) = P(x - S_2S_1x). \tag{8}$$

The remainder of this section is devoted to providing smoothness and convexity properties of the envelope function under Assumption 3.2.

### 3.1 Basic Properties of the Envelope Function

The following two results are special cases and direct corollaries of a more general result in Theorem 3.1, to be presented later. Proofs are therefore omitted.

**Proposition 3.2** *Suppose that Assumption 3.2 holds. Then, the gradient of  $F$  is 2-Lipschitz continuous. That is,  $\nabla F$  satisfies*

$$\|\nabla F(x) - \nabla F(y)\| \leq 2\|x - y\|$$

for all  $x, y \in \mathbb{R}^n$ .

**Proposition 3.3** *Suppose that Assumption 3.2 holds and that  $P$ , that defines the linear part of  $S_1$ , is positive semidefinite. Then,  $F$  is convex.*

If  $P$  is positive semidefinite, then the envelope function  $F$  is convex and differentiable with a Lipschitz continuous gradient. This implies, e.g., that all stationary points are minimizers. If  $P$  is positive definite we know from Proposition 3.1 that the set of stationary points coincides with the fixed-point set of  $S = S_2S_1$ . Therefore, a fixed-point to  $S_2S_1$  can be found by minimizing the smooth convex envelope function  $F$ .

### 3.2 Finer Properties of the Envelope Function

In this section, we establish sharp upper and lower bounds for the envelope function (7). These results use stronger assumptions on  $S_2$  than nonexpansiveness, namely that  $S_2$  is  $\alpha$ -averaged and  $\beta$ -negatively averaged:

**Assumption 3.3** The operator  $S_2$  is  $\alpha$ -averaged and  $\beta$ -negatively averaged with  $\alpha \in ]0, 1]$  and  $\beta \in ]0, 1]$ .

Before we proceed, we state a result on how averaged and negatively averaged gradient operators can equivalently be characterized. The result is proven in Appendix A.

**Lemma 3.1** Assume that  $f$  is differentiable. Then,  $\nabla f$  is  $\alpha$ -averaged with  $\alpha \in ]0, 1]$  and  $\beta$ -negatively averaged with  $\beta \in ]0, 1]$  if and only if

$$-\frac{2\alpha-1}{2} \|x - y\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{2\beta-1}{2} \|x - y\|^2 \tag{9}$$

holds for all  $x, y \in \mathbb{R}^n$ , which holds if and only if

$$-(2\alpha - 1)\|x - y\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq (2\beta - 1)\|x - y\|^2 \tag{10}$$

holds for all  $x, y \in \mathbb{R}^n$ .

These properties relate to smoothness and strong convexity properties of  $f$ . More precisely, they imply that  $f$  is  $\max((2\alpha - 1), (2\beta - 1))$ -smooth and, if  $\alpha > \frac{1}{2}$ ,  $(2\alpha - 1)$ -strongly convex. With this interpretation in mind, we state the main theorem.

**Theorem 3.1** Suppose that Assumptions 3.2 and 3.3 hold. Further, let  $\delta_\alpha = 2\alpha - 1$  and  $\delta_\beta = 2\beta - 1$ . Then, the envelope function  $F$  in (7) satisfies

$$F(x) - F(y) - \langle \nabla F(y), x - y \rangle \geq \frac{1}{2} \left\langle (P - \delta_\beta P^2) (x - y), x - y \right\rangle$$

and

$$F(x) - F(y) - \langle \nabla F(y), x - y \rangle \leq \frac{1}{2} \left\langle (P + \delta_\alpha P^2) (x - y), x - y \right\rangle$$

for all  $x, y \in \mathbb{R}^n$ . Furthermore, the bounds are tight.

A proof to this result is found in “Appendix B”.

Utilizing connections established in Sect. 2.2.3, we next derive different properties of the envelope function. Especially, we provide conditions under which the envelope function is convex and strongly convex.

**Corollary 3.1** Suppose that the assumptions of Theorem 3.1 hold and that  $P$  is positive semidefinite. Then,

$$\frac{1}{2} \|x - y\|_{P - \delta_\beta P^2}^2 \leq F(x) - F(y) - \langle \nabla F(y), x - y \rangle \leq \frac{1}{2} \|x - y\|_{P + \delta_\alpha P^2}^2$$



and  $F$  is convex and  $I$ -smooth w.r.t.  $\|\cdot\|_{P+\delta_\alpha P^2}$ . If in addition  $P$  is positive definite and either of the following holds:

- (i)  $P$  is contractive,
- (ii)  $\beta \in ]0, 1[$ , i.e.,  $\delta_\beta \in ]-1, 1[$ ,

then  $F$  is  $I$ -strongly convex w.r.t.  $\|\cdot\|_{P-\delta_\beta P^2}$  and  $I$ -smooth w.r.t.  $\|\cdot\|_{P+\delta_\alpha P^2}$ .

*Proof* The results follow from Theorem 3.1, the definition of (strong) convexity, and by utilizing Lemma D.3 in “Appendix D” to show that the smallest eigenvalue of  $P - \delta_\beta P^2$  is nonnegative and positive, respectively.  $\square$

Less sharp, but unscaled, versions of these bounds can easily be obtained from Theorem 3.1.

**Corollary 3.2** *Suppose that the assumptions of Theorem 3.1 hold. Then,*

$$\frac{\beta_l}{2} \|x - y\|^2 \leq F(x) - F(y) - \langle \nabla F(y), x - y \rangle \leq \frac{\beta_u}{2} \|x - y\|^2,$$

where  $\beta_l = \lambda_{\min}(P - \delta_\beta P^2)$  and  $\beta_u = \lambda_{\max}(P + \delta_\alpha P^2)$ .

Values of  $\beta_l$  and  $\beta_u$  for different assumptions on  $P$ ,  $\delta_\alpha$  and  $\delta_\beta$  can be obtained from Lemma D.3 in “Appendix D”.

The results in Theorem 3.1 and its corollaries are stated for  $\alpha$ -averaged and  $\beta$ -negatively averaged operators  $S_2 = \nabla f_2$ . Using Lemmas 3.1 and D.2, we conclude that  $\delta$ -contractive operators are  $\alpha$ -averaged and  $\beta$ -negatively averaged with  $\alpha$  and  $\beta$  satisfying  $\delta = \delta_\alpha = \delta_\beta$ . This gives the following result.

**Proposition 3.4** *Suppose that Assumption 3.2 holds and that  $S_2$  is  $\delta$ -Lipschitz continuous with  $\delta \in [0, 1]$ . Then, all results in this section hold with  $\delta_\beta$  and  $\delta_\alpha$  replaced by  $\delta$ .*

If instead  $S_2 = \nabla f_2$  is  $\frac{1}{\delta}$ -cocoercive, it can be shown (see [28, Definition 4.4] and [30, Theorem 2.1.5]) that

$$0 \leq f_2(x) - f_2(y) - \langle \nabla f_2(y), x - y \rangle \leq \frac{\delta}{2} \|x - y\|^2.$$

In view of Lemma 3.1, we can state the following result.

**Proposition 3.5** *Suppose that Assumption 3.2 holds and that  $S_2$  is  $\frac{1}{\delta}$ -cocoercive with  $\delta \in ]0, 1]$ . Then, all results in this section hold with  $\delta_\beta = \delta$  and  $\delta_\alpha = 0$ .*

### 3.3 Majorization–Minimization Interpretation of Averaged Iteration

As noted in [22,24], the forward–backward and Douglas–Rachford splitting methods are variable metric gradient methods applied to their respective envelope functions. In our setting, with  $S_1$  being affine, they reduce to being fixed-metric scaled gradient methods. In this section, we provide a different interpretation. We show that a step in the basic iteration is obtained by performing majorization–minimization on the

envelope. The majorizing function is a closely related to the upper bound provided in Corollary 3.1.

The interpretation is valid under the assumption that  $P$  is positive definite, besides being nonexpansive. This implies that the envelope is convex, see Corollary 3.1. It is straightforward to verify that  $P + \delta_\alpha P^2 \leq (1 + \delta_\alpha)P$ . Therefore, we can construct the following more conservative upper bound to the envelope, compared to Corollary 3.1:

$$F(x) \leq F(y) + \langle \nabla F(y), x - y \rangle + \frac{1+\delta_\alpha}{2} \|x - y\|_P^2. \tag{11}$$

Minimizing this majorizer, evaluated at  $y = x^k$ , in every iteration  $k$  gives

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_x \left\{ F(x^k) + \langle \nabla F(x^k), x - x^k \rangle + \frac{1 + \delta_\alpha}{2} \|x - x^k\|_P^2 \right\} \\ &= x^k - \frac{1}{1 + \delta_\alpha} P^{-1} \nabla F(x^k) \\ &= x^k - \frac{1}{1 + \delta_\alpha} P^{-1} P(S_2 S_1 x^k - x^k) \\ &= x^k - \frac{1}{1 + \delta_\alpha} (S_2 S_1 x^k - x^k) \\ &= \left( 1 - \frac{1}{1 + \delta_\alpha} \right) x^k + \frac{1}{1 + \delta_\alpha} S_2 S_1 x^k, \end{aligned}$$

which is the basic method with  $\frac{1}{1+\delta_\alpha}$ -averaging. It is well known that the gradient method converges with step-length  $\alpha \in ]0, \frac{2}{L}[$ , where  $L$  is a Lipschitz constant. In this case, the upper bound (11) guarantees a Lipschitz constant to  $\nabla F$  of  $L = 1 + \delta_\alpha$  in the  $\|\cdot\|_P$ -norm, see Lemma D.2. Selecting a step-length within the allowed range yields an averaged iteration with  $\frac{1}{1+\delta_\alpha}$  replaced by  $\alpha \in ]0, \frac{2}{1+\delta_\alpha}[$ .

The upper bound (11) used to arrive at the averaged iteration is not sharp. Using instead the sharp majorizer from Corollary 3.1, yields the following algorithm:

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_x \left\{ F(x^k) + \langle \nabla F(x^k), x - x^k \rangle + \frac{1}{2} \|x - x^k\|_{P+\delta_\alpha P^2}^2 \right\} \\ &= x^k - (\operatorname{Id} + \delta_\alpha P)^{-1} P^{-1} \nabla F(x^k) \\ &= x^k - (\operatorname{Id} + \delta_\alpha P)^{-1} P^{-1} P(S_2 S_1 x^k - x^k) \\ &= x^k - (\operatorname{Id} + \delta_\alpha P)^{-1} (S_2 S_1 x^k - x^k) \\ &= (\operatorname{Id} - (\operatorname{Id} + \delta_\alpha P)^{-1}) x^k + (\operatorname{Id} + \delta_\alpha P)^{-1} S_2 S_1 x^k. \end{aligned}$$

This differs from the basic averaged iteration in that  $(1 + \delta_\alpha)^{-1} \operatorname{Id}$  in the basic method is replaced by  $(\operatorname{Id} + \delta_\alpha P)^{-1}$ . The drawback of using this tighter majorizer is that the iterations become more expensive.

None of these methods is probably the most efficient way to find a stationary point of the envelope function (or equivalently a fixed-point to  $S_2 S_1$ ). At least in the convex setting (for the envelope), there are numerous alternative methods that can minimize

smooth functions such as truncated Newton methods, quasi-Newton methods, and nonlinear conjugate gradient methods. See [31] for an overview of such methods and [22,23] for some of these methods applied to the forward–backward envelope. Evaluating which ones that are most efficient and devising new methods to improve performance is outside the scope of this paper.

### 4 Special Cases

In this section, we show that our envelope in (5) has four known special cases, namely the Moreau envelope [32], the forward–backward envelope [22,23], the Douglas–Rachford envelope [24], and the ADMM envelope [27] (which is a special case of the Douglas–Rachford envelope).

We also show that our envelope bounds for  $S_1 = \nabla f_1$  being affine coincide with or sharpen corresponding results in the literature for the special cases.

#### 4.1 Algorithm Building Blocks

Before we present the special cases, we introduce some functions, whose gradients are operators that are used in the respective underlying methods. Most importantly, we will introduce a function whose gradient is the proximal operator:

$$\text{prox}_{\gamma f}(z) := \underset{x}{\operatorname{argmin}}\{f(x) + \frac{1}{2\gamma}\|x - z\|^2\},$$

where  $\gamma > 0$  is a parameter.

**Proposition 4.1** *Suppose that  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  is proper, closed, and convex and that  $\gamma > 0$ . The proximal operator  $\text{prox}_{\gamma f}$  then satisfies*

$$\text{prox}_{\gamma f} = \nabla r_{\gamma f}^*,$$

where  $r_{\gamma f}^*$  is the conjugate of

$$r_{\gamma f}(x) := \gamma f(x) + \frac{1}{2}\|x\|^2. \tag{12}$$

The reflected proximal operator

$$R_{\gamma f} := 2\text{prox}_{\gamma f} - \text{Id} \tag{13}$$

satisfies  $R_{\gamma f} = \nabla p_{\gamma f}$ , where

$$p_{\gamma f} := 2r_{\gamma f}^* - \frac{1}{2}\|\cdot\|^2. \tag{14}$$

This proximal map interpretation is from [33, Theorems 31.5, 16.4] and implies that the proximal operator is the gradient of a convex function. The reflected proximal operator interpretation follows trivially from the prox interpretation.

The other algorithm building block that is used in the considered algorithms is the gradient step. The gradient step operator is the gradient of the function  $\frac{1}{2}\|x\|^2 - \gamma f(x)$ , i.e.,:

$$(x - \gamma \nabla f(x)) = \nabla \left( \frac{1}{2}\|x\|^2 - \gamma f(x) \right).$$

## 4.2 The Proximal Point Algorithm

The proximal point algorithm solves problems of the form

$$\text{minimize } f(x),$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  is proper, closed, and convex.

The algorithm repeatedly applies the proximal operator of  $f$  and is given by

$$x^{k+1} = \text{prox}_{\gamma f}(x^k), \quad (15)$$

where  $\gamma > 0$  is a parameter. This algorithm is mostly of conceptual interest since it is often as computationally demanding to evaluate the prox as to minimize the function  $f$  itself.

Its envelope function, which is called the Moreau envelope [32], is a scaled version of the envelope  $F$  in (7). The scaling factor is  $\gamma^{-1}$  and the Moreau envelope  $f^\gamma$  is obtained by letting  $S_1 x = \nabla f_1(x) = x$ , i.e.,  $P = \text{Id}$  and  $q = 0$ , and  $f_2 = r_{\gamma f}^*$  in (7), where  $r_{\gamma f}$  is defined in (12):

$$f^\gamma(x) = \gamma^{-1} F(x) = \gamma^{-1} \left( \frac{1}{2}\|x\|^2 - r_{\gamma f}^*(x) \right). \quad (16)$$

Its gradient satisfies

$$\nabla f^\gamma(x) = \gamma^{-1} (x - \text{prox}_{\gamma f}(x)).$$

The following properties of the Moreau envelope follow directly from Corollary 3.2 and Proposition 3.5 since the proximal operator is 1-cocoercive (see Remark 2.4 and [28, Proposition 12.27]).

**Proposition 4.2** *The Moreau envelope  $f^\gamma$  in (16) is differentiable and convex and  $\nabla f^\gamma$  is  $\gamma^{-1}$ -Lipschitz continuous.*

This coincides with previously known properties of the Moreau envelope, see [28, Chapter 12].

## 4.3 Forward–Backward Splitting

Forward–backward splitting solves problems of the form

$$\text{minimize } f(x) + g(x), \quad (17)$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex with an  $L$ -Lipschitz (or equivalently  $\frac{1}{L}$ -cocoercive) gradient, and  $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  is proper, closed, and convex.

The algorithm performs a forward step followed by a backward step, and is given by

$$x^{k+1} = \text{prox}_{\gamma g}(\text{Id} - \gamma \nabla f)x^k, \tag{18}$$

where  $\gamma \in ]0, \frac{2}{L}[$  is a parameter.

The envelope function, which is called the forward–backward envelope [22,23], is a scaled version of the envelope  $F$  in (5) and applies when  $f$  is twice continuously differentiable. The scaling factor is  $\gamma^{-1}$  and the forward–backward envelope is obtained by letting  $f_1 = \frac{1}{2}\|\cdot\|^2 - \gamma f$  and  $f_2 = r_{\gamma g}^*$  in (5), where  $r_{\gamma g}$  is defined in (12). The resulting forward–backward envelope function is

$$F_\gamma^{\text{FB}}(x) = \gamma^{-1} \left( \langle x - \gamma \nabla f(x), x \rangle - \left( \frac{1}{2}\|x\|^2 - \gamma f(x) \right) - r_{\gamma g}^*(x - \gamma \nabla f(x)) \right).$$

The gradient of this function is

$$\begin{aligned} \nabla F_\gamma^{\text{FB}}(x) &= \gamma^{-1} ((\text{Id} - \gamma \nabla^2 f(x))x + (x - \gamma \nabla f(x)) - (x - \gamma \nabla f(x)) \\ &\quad - (\text{Id} - \gamma \nabla^2 f(x))\text{prox}_{\gamma g}(x - \gamma \nabla f(x))) \\ &= \gamma^{-1} (\text{Id} - \gamma \nabla^2 f(x)) (x - \text{prox}_{\gamma g}(x - \gamma \nabla f(x))), \end{aligned}$$

which coincides with the gradient in [22,23]. As described in [22,23], the stationary points of the envelope coincide with the fixed-points of the mapping  $\text{prox}_{\gamma g}(x - \gamma \nabla f(x))$  if  $(\text{Id} - \gamma \nabla^2 f(x))$  is nonsingular.

### 4.3.1 $S_1$ Affine

We provide properties of the forward–backward envelope in the more restrictive setting of  $S_1 = \nabla f_1 = (\text{Id} - \gamma \nabla f)$  being affine. This applies when  $f$  is a convex quadratic,  $f(x) = \frac{1}{2}\langle Hx, x \rangle + \langle h, x \rangle$  with  $H \in \mathbb{R}^{n \times n}$  positive semidefinite and  $h \in \mathbb{R}^n$ . Then,  $S_1x = Px + q$  with  $P = (\text{Id} - \gamma H)$  and  $q = -\gamma h$ .

In this setting, the following result follows immediately from Corollary 3.1 and Proposition 3.5 (where Proposition 3.5 is invoked since  $S_2 = \text{prox}_{\gamma g}$  is 1-cocoercive, see Remark 2.4 and [28, Proposition 12.27]).

**Proposition 4.3** *Assume that  $f(x) = \frac{1}{2}\langle Hx, x \rangle + \langle h, x \rangle$  and  $\gamma \in ]0, \frac{1}{L}[$ , where  $L = \lambda_{\max}(H)$ . Then, the forward–backward envelope  $F_\gamma^{\text{FB}}$  satisfies*

$$\frac{1}{2\gamma}\|x - y\|_{P-P^2}^2 \leq F_\gamma^{\text{FB}}(x) - F_\gamma^{\text{FB}}(y) - \langle \nabla F_\gamma^{\text{FB}}(y), x - y \rangle \leq \frac{1}{2\gamma}\|x - y\|_P^2$$

for all  $x, y \in \mathbb{R}^n$ , where  $P = (\text{Id} - \gamma H)$  is positive definite. If in addition  $\lambda_{\min}(H) = m > 0$ , then  $P - P^2$  is positive definite and  $F_\gamma^{\text{FB}}$  is  $\gamma^{-1}$ -strongly convex w.r.t.  $\|\cdot\|_{P-P^2}$ .

Less tight bounds for the forward–backward envelope are provided next. These follow immediately from the above and Lemma D.3.

**Proposition 4.4** *Assume that  $f(x) = \frac{1}{2}\langle Hx, x \rangle + \langle h, x \rangle$ , that  $\gamma \in ]0, \frac{1}{L}[$  where  $L = \lambda_{\max}(H)$ , and that  $m = \lambda_{\min}(H) \geq 0$ . Then, the forward–backward envelope  $F_\gamma^{\text{FB}}$  is  $\gamma^{-1}(1 - \gamma m)$ -smooth and  $\min((1 - \gamma m)m, (1 - \gamma L)L)$ -strongly convex (both w.r.t. to the induced norm  $\| \cdot \|$ ).*

This result is a less tight version of Proposition 4.3, but is a slight improvement of the corresponding result in [22, Theorem 2.3]. The strong convexity moduli are the same, but our smoothness constant is a factor two smaller.

### 4.4 Douglas–Rachford Splitting

Douglas–Rachford splitting solves problems of the form

$$\text{minimize } f(x) + g(x), \tag{19}$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  are proper, closed, and convex functions.

The algorithm performs two reflection steps (13), then an averaging:

$$z^{k+1} = (1 - \alpha)z^k + \alpha R_{\gamma g} R_{\gamma f} z^k, \tag{20}$$

where  $\gamma > 0$  and  $\alpha \in ]0, 1[$  are parameters. The objective is to find a fixed-point  $\bar{z}$  to  $R_{\gamma g} R_{\gamma f}$ , from which a solution to (19) can be computed as  $\text{prox}_{\gamma f}(\bar{z})$ , see [28, Proposition 25.1].

The envelope function in [24], which is called the Douglas–Rachford envelope, is a scaled version of the basic envelope function  $F$  in (5) and applies when  $f$  is twice continuously differentiable and  $\nabla f$  is Lipschitz continuous. The scaling factor is  $(2\gamma)^{-1}$  and the Douglas–Rachford envelope is obtained by, in (5), letting  $f_1 = p_{\gamma f}$  with gradient  $\nabla f_1 = S_1 = R_{\gamma f}$  and  $f_2 = p_{\gamma g}$ , where  $p_{\gamma g}$  is defined in (14). The Douglas–Rachford envelope function becomes

$$F_\gamma^{\text{DR}}(z) = (2\gamma)^{-1} (\langle R_{\gamma f}(z), z \rangle - p_{\gamma f}(z) - p_{\gamma g}(R_{\gamma f}z)). \tag{21}$$

The gradient of this function is

$$\begin{aligned} \nabla F_\gamma^{\text{DR}}(z) &= (2\gamma)^{-1} (\nabla R_{\gamma f}(z)z + R_{\gamma f} - R_{\gamma f} - \nabla R_{\gamma f}(z)R_{\gamma g}(R_{\gamma f}(z))) \\ &= (2\gamma)^{-1} \nabla R_{\gamma f}(z)(z - R_{\gamma g}R_{\gamma f}(z)), \end{aligned}$$

which coincides with the gradient in [24] since  $\nabla R_{\gamma f} = 2\nabla \text{prox}_{\gamma f} - \text{Id}$  and

$$\begin{aligned} z - R_{\gamma g}R_{\gamma f}z &= z - 2\text{prox}_{\gamma g}(2\text{prox}_{\gamma f}(z) - z) + 2\text{prox}_{\gamma f}(z) - z \\ &= 2(\text{prox}_{\gamma f}(z) - \text{prox}_{\gamma g}(2\text{prox}_{\gamma f}(z) - z)). \end{aligned}$$

As described in [24], the stationary points of the envelope coincide with the fixed-points of  $R_{\gamma g}R_{\gamma f}$  if  $\nabla R_{\gamma f}$  is nonsingular.

#### 4.4.1 $S_1$ Affine

We state properties of the Douglas–Rachford envelope in the more restrictive setting of  $S_1 = R_{\gamma f}$  being affine. This is obtained for convex quadratic  $f$ :

$$f(x) = \frac{1}{2}\langle Hx, x \rangle + \langle h, x \rangle,$$

where  $H$  is positive semidefinite. The operator  $S_1$  becomes

$$S_1(z) = R_{\gamma f}(z) = 2(\text{Id} + \gamma H)^{-1}(z - \gamma h) - z,$$

which confirms that it is affine. We implicitly define  $P$  and  $q$  through the relation  $S_1 = R_{\gamma f} = P(\cdot) + q$ , and note that they are given by the expressions  $P = 2(\text{Id} + \gamma H)^{-1} - \text{Id}$  and  $q = -2\gamma(\text{Id} + \gamma H)^{-1}h$ , respectively.

In this setting, the following result follows immediately from Corollary 3.1 since  $S_2 = R_{\gamma g}$  is nonexpansive (1-averaged and 1-negatively averaged).

**Proposition 4.5** *Assume that  $f(x) = \frac{1}{2}\langle Hx, x \rangle + \langle h, x \rangle$  and  $\gamma \in ]0, \frac{1}{L}[$ , where  $L = \lambda_{\max}(H)$ . Then, the Douglas–Rachford envelope  $F_{\gamma}^{\text{DR}}$  satisfies*

$$\frac{1}{4\gamma} \|z - y\|_{P-P^2}^2 \leq F_{\gamma}^{\text{DR}}(z) - F_{\gamma}^{\text{DR}}(y) - \langle \nabla F_{\gamma}^{\text{DR}}(y), z - y \rangle \leq \frac{1}{4\gamma} \|z - y\|_{P+P^2}^2$$

for all  $y, z \in \mathbb{R}^n$ , where  $P = 2(\text{Id} + \gamma H)^{-1} - \text{Id}$  is positive definite. If in addition  $\lambda_{\min}(H) = m > 0$ , then  $P - P^2$  is positive definite and  $F_{\gamma}^{\text{DR}}$  is  $(2\gamma)^{-1}$ -strongly convex w.r.t.  $\|\cdot\|_{P-P^2}$ .

The following less tight characterization of the Douglas–Rachford envelope follows from the above and Lemma D.3.

**Proposition 4.6** *Assume that  $f(x) = \frac{1}{2}\langle Hx, x \rangle + \langle h, x \rangle$ , that  $\gamma \in ]0, \frac{1}{L}[$ , where  $L = \lambda_{\max}(H)$ , and that  $m = \lambda_{\min}(H) \geq 0$ . Then, the Douglas–Rachford envelope  $F_{\gamma}^{\text{DR}}$  is  $\frac{1-\gamma m}{(1+\gamma m)^2} \gamma^{-1}$ -smooth and  $\min\left(\frac{(1-\gamma m)m}{(1+\gamma m)^2}, \frac{(1-\gamma L)L}{(1+\gamma L)^2}\right)$ -strongly convex.*

This result is more conservative than the one in Proposition 4.5, but improves on [24, Theorem 2]. The strong convexity modulus coincides with the corresponding one in [24, Theorem 2]. The smoothness constant is  $\frac{1}{1+\gamma m}$  times that in [24, Theorem 2], i.e., it is slightly smaller.

### 4.5 ADMM

The alternating direction method of multipliers (ADMM) solves problems of the form (19). It is well known [34] that ADMM can be interpreted as Douglas–Rachford applied to the dual of (19), namely to

$$\text{minimize } f^*(\mu) + g^*(-\mu). \tag{22}$$

So the algorithm is given by

$$v^{k+1} = (1 - \alpha)v^k + \alpha R_{\rho(g^* \circ -\text{Id})} R_{\rho f} v^k, \tag{23}$$

where  $\rho > 0$  is a parameter,  $R_{\rho f}$  is the reflected proximal operator (13), and  $(g^* \circ -\text{Id})$  is the composition that satisfies  $(g^* \circ -\text{Id})(\mu) = g^*(-\mu)$ .

In accordance with the Douglas–Rachford envelope (21), the ADMM envelope is

$$F_{\rho}^{\text{ADMM}}(v) = (2\rho)^{-1} \left( \langle R_{\rho f^*}(v), v \rangle - p_{\rho f^*}^2(v) - p_{\rho(g^* \circ -\text{Id})}^2(R_{\rho f^*} v) \right) \tag{24}$$

and its gradient becomes

$$\nabla F_{\rho}^{\text{ADMM}}(v) = (2\rho)^{-1} \nabla R_{\rho f^*}(v)(v - R_{\rho(g^* \circ -\text{Id})} R_{\rho f^*}(v)).$$

This envelope function has been utilized in [27] to accelerate performance of ADMM. In this section, we will augment the analysis in [27] by relating the ADMM algorithm and its envelope function to the Douglas–Rachford counterparts. To do so, we need the following result which is proven in ‘‘Appendix C’’.

**Lemma 4.1** *Let  $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, closed, and convex and let  $\rho > 0$ . Then,*

$$\begin{aligned} R_{\rho g^*}(x) &= -\rho R_{\rho^{-1}g}(\rho^{-1}x), \\ R_{\rho(g^* \circ -\text{Id})}(x) &= \rho R_{\rho^{-1}g}(-\rho^{-1}x), \\ p_{\rho(g^* \circ -\text{Id})}(y) &= -\rho^2 p_{\rho^{-1}g}(-\rho^{-1}y), \end{aligned}$$

where  $R_{\rho g}$  is defined in (13) and  $p_{\rho g}$  is defined in (14).

Before we state the result, we show that the  $z^k$  sequence in (primal) Douglas–Rachford (20) and the  $v^k$  sequence in ADMM (i.e., dual Douglas–Rachford) in (23) differ by a factor only. This is well known [35], but the relation is stated next with a simple proof.

**Proposition 4.7** *Assume that  $\rho > 0$  and  $\gamma > 0$  satisfy  $\rho^{-1} = \gamma$ , and that  $z^0 = \rho^{-1}v^0$ . Then  $z^k = \rho^{-1}v^k$  for all  $k \geq 1$ , where  $\{z^k\}$  is the primal Douglas–Rachford sequence defined in (20) and the  $\{v^k\}$  is the ADMM sequence is defined in (23).*

*Proof* Lemma 4.1 implies that

$$\begin{aligned} v^{k+1} &= (1 - \alpha)v^k + \alpha R_{\rho(g^* \circ -\text{Id})} R_{\rho f^*} v^k \\ &= (1 - \alpha)v^k + \alpha \rho R_{\rho^{-1}g}(-\rho^{-1}(-\rho R_{\rho^{-1}f}(\rho^{-1}v^k))) \\ &= (1 - \alpha)v^k + \alpha \rho R_{\rho^{-1}g}(R_{\rho^{-1}f}(\rho^{-1}v^k)). \end{aligned}$$



Multiply by  $\rho^{-1}$ , let  $z^k = \rho^{-1}v^k$ , and identify  $\gamma = \rho^{-1}$  to get

$$z^{k+1} = (1 - \alpha)z^k + \alpha R_{\gamma g}(R_{\gamma f}(z^k)).$$

This concludes the proof. □

There is also a tight relationship between the ADMM and Douglas–Rachford envelopes. Essentially, they have opposite signs.

**Proposition 4.8** *Assume that  $\rho > 0$  and  $\gamma > 0$  satisfy  $\rho = \gamma^{-1}$  and that  $z = \rho^{-1}v = \gamma v$ . Then,*

$$F_{\rho}^{\text{ADMM}}(v) = -F_{\gamma}^{\text{DR}}(z).$$

*Proof* Using Lemma 4.1 several times,  $\gamma = \rho^{-1}$ , and  $z = \rho^{-1}v$ , we conclude that

$$\begin{aligned} F_{\rho}^{\text{ADMM}}(v) &= (2\rho)^{-1} (\langle R_{\rho f^*}(v), v \rangle - p_{\rho f^*}(v) - p_{\rho(g \circ \text{Id})}(R_{\rho f^*}(v))) \\ &= (2\rho)^{-1} \left( -\rho \langle R_{\rho^{-1}f}(\rho^{-1}v), v \rangle + \rho^2 p_{\rho^{-1}(f \circ \text{Id})}(-\rho^{-1}v) \right. \\ &\quad \left. + \rho^2 p_{\rho^{-1}g}(-\rho^{-1}(-\rho R_{\rho^{-1}f}(\rho^{-1}v))) \right) \\ &= -\frac{\rho}{2} \left( \langle R_{\rho^{-1}f}(\rho^{-1}v), \rho^{-1}v \rangle - p_{\rho^{-1}f}(\rho^{-1}v) + p_{\rho^{-1}g}(R_{\rho^{-1}f}(\rho^{-1}v)) \right) \\ &= -(2\gamma)^{-1} (\langle R_{\gamma f}(z), z \rangle - p_{\gamma f}(z) + p_{\gamma g}(R_{\gamma f}(z))) \\ &= -F_{\gamma}^{\text{DR}}(z). \end{aligned}$$

This concludes the proof. □

This result implies that the ADMM envelope is concave when the DR envelope is convex, and vice versa. We know from Sect. 4.4 that the operator  $S_1 = R_{\rho f^*}$  is affine when the conjugate  $f^*$  is quadratic. This holds true if

$$f(x) = \begin{cases} \frac{1}{2} \langle Hx, x \rangle + \langle h, x \rangle, & \text{if } Ax = b, \\ \infty, & \text{else,} \end{cases}$$

and  $H$  is positive definite on the nullspace of  $A$ . From Propositions 4.5 and 4.6, we conclude that, for an appropriate choice of  $\rho$ , the ADMM envelope is convex, which implies that the Douglas–Rachford envelope is concave.

*Remark 4.1* The standard ADMM formulation is applied to solve problems of the form

$$\begin{aligned} &\text{minimize} && \hat{f}(x) + \hat{g}(z) \\ &\text{subject to} && Ax + Bz = c. \end{aligned}$$

Using infimal post-compositions, also called image functions, the dual of this is on the form (22), see, e.g., [36, Appendix B], which is a longer version of [37], for details. Therefore also this setting is implicitly considered.

## 5 Conclusions

We have presented an envelope function that unifies the Moreau envelope, the forward–backward envelope, the Douglas–Rachford envelope, and the ADMM envelope. We have provided quadratic upper and lower bounds for the envelope that coincide with or improve on corresponding results in the literature for the special cases. We have also provided a novel interpretation of the underlying algorithms as being majorization–minimization algorithms applied to their respective envelopes. Finally, we have shown how the ADMM and DR envelopes relate to each other.

**Acknowledgements** Pontus Giselsson and Mattias Fält are financially supported by the Swedish Foundation for Strategic Research and members of the LCCC Linneaus Center at Lund University. Pontus Giselsson is also financed by the Swedish Research Council. The reviewers are gratefully acknowledged for useful comments that have considerably improved the paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendices

### Appendix A: Proof to Lemma 3.1

The operator  $\nabla f$  is  $\alpha$ -averaged if and only if  $\nabla f = (1 - \alpha)\text{Id} + \alpha R$  for some nonexpansive operator  $R$ . Therefore,  $\nabla f$  is  $\alpha$ -averaged if and only if  $\nabla f - (1 - \alpha)\text{Id}$  is  $\alpha$ -Lipschitz continuous, since  $\nabla f - (1 - \alpha)\text{Id} = \alpha R$ . Letting  $g := f - \frac{1-\alpha}{2} \|\cdot\|^2$ , we get  $\nabla g = \alpha R$ . Therefore  $\nabla g$  is  $\alpha$ -Lipschitz. According to Lemma D.2 this is equivalent to that

$$|g(x) - g(y) - \langle \nabla g(y), x - y \rangle| \leq \frac{\alpha}{2} \|x - y\|^2$$

or equivalently

$$|f(x) - f(y) - \langle \nabla f(y), x - y \rangle - \frac{1-\alpha}{2} \|x - y\|^2| \leq \frac{\alpha}{2} \|x - y\|^2,$$

which is equivalent to

$$-\frac{2\alpha-1}{2} \|x - y\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{1}{2} \|x - y\|^2. \quad (25)$$

The  $\beta$ -negative averagedness is defined as that  $-\nabla f$  is  $\beta$ -averaged. Similar arguments as the above give that  $\nabla f$  is  $\beta$ -negatively averaged if and only if

$$-\frac{1}{2} \|x - y\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{2\beta-1}{2} \|x - y\|^2. \quad (26)$$

Now, the upper bound in (25) and the lower bound in (26) are redundant and we arrive at

$$-\frac{2\alpha-1}{2}\|x-y\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x-y \rangle \leq \frac{2\beta-1}{2}\|x-y\|^2$$

to prove the first equivalence. The second equivalence follows from Lemma D.1.

### Appendix B: Proof to Theorem 3.1

First, we establish that

$$-\delta_\alpha\|x-y\|_{P^2}^2 \leq \langle P\nabla f_2(Px+q) - P\nabla f_2(Py+q), x-y \rangle \leq \delta_\beta\|x-y\|_{P^2}^2. \tag{27}$$

We have

$$\begin{aligned} &\langle P\nabla f_2(Px+q) - P\nabla f_2(Py+q), x-y \rangle \\ &= \langle \nabla f_2(Px+q) - \nabla f_2(Py+q), P(x-y) \rangle \\ &= \langle \nabla f_2(Px+q) - \nabla f_2(Py+q), (Px+q) - (Py+q) \rangle. \end{aligned}$$

This implies that

$$\begin{aligned} -(2\alpha-1)\|x-y\|_{P^2}^2 &= -(2\alpha-1)\|(Px+q) - (Py+q)\|^2 \\ &\leq \langle P\nabla f_2(Px+q) - P\nabla f_2(Py+q), x-y \rangle \\ &\leq (2\beta-1)\|(Px+q) - (Py+q)\|^2 \\ &= (2\beta-1)\|x-y\|_{P^2}^2, \end{aligned}$$

where Lemma 3.1 is used in the inequalities. Recalling that  $\delta_\alpha = 2\alpha - 1$  and  $\delta_\beta = 2\beta - 1$ , this shows that (27) holds. In addition, for any  $\delta \in \mathbb{R}$ , we have

$$\begin{aligned} \langle \nabla F(x) - \nabla F(y), x-y \rangle &= \langle P(x - \nabla f_2 \nabla f_1(x)) - P(x - \nabla f_2 \nabla f_1(y)), x-y \rangle \\ &= \langle P(x-y), x-y \rangle \\ &\quad - \langle P\nabla f_2(Px+q) - P\nabla f_2(Py+q), x-y \rangle \\ &= \langle (P - \delta P^2)(x-y), x-y \rangle + \delta\|x-y\|_{P^2}^2 \\ &\quad - \langle P\nabla f_2(Px+q) - P\nabla f_2(Py+q), x-y \rangle. \tag{28} \end{aligned}$$

Let  $\delta = -\delta_\alpha$ , then (28) and (27) imply

$$\langle \nabla F(x) - \nabla F(y), x-y \rangle \leq \langle (P + \delta_\alpha P^2)(x-y), x-y \rangle.$$

Let  $\delta = \delta_\beta$ , then (28) and (27) imply

$$\langle \nabla F(x) - \nabla F(y), x-y \rangle \geq \langle (P - \delta_\beta P^2)(x-y), x-y \rangle.$$

Applying Lemma D.1 in “Appendix D” gives the result.

Next, we show that the bounds are sharp. The obtained inequality implies through Lemmas D.1 and D.2 that  $\nabla F$  is Lipschitz continuous. Hence, by Rademacher’s Theorem, it is differentiable almost everywhere, i.e.,  $\partial^2 F$  is unique almost everywhere. Using [38, Proposition 2.6.2d], we can conclude from the upper and lower bounds, Lemmas D.1, and D.2 that  $P - \delta_\beta P^2 \leq \partial^2 F(x) \leq P + \delta_\alpha P^2$ . Now, let us select a point where  $\partial^2 F(x) = \{\nabla^2 F(x)\}$ . The Hessian satisfies

$$\nabla^2 F(x) = \nabla(Px - P\nabla f_2(Px + q)) = P - P^2\nabla^2 f_2(Px + q).$$

Now, select a function  $f_2$  with  $\beta$ -negatively averaged gradient  $\nabla f_2$  such that its Hessian at  $Px + q$  satisfies  $\nabla^2 f_2(Px + q) = -\delta_\beta \text{Id}$  (e.g., by letting  $\nabla f_2(x) = -\delta_\beta x$ , which is  $\beta$ -negatively averaged). Then,  $\nabla^2 F(x) = P + \delta_\beta P^2$ , which shows that the lower bound is tight. Similar arguments show that the upper bound can be attained.

### Appendix C: Proof to Lemma 4.1

Using the Moreau decomposition [28, Theorem 14.3]

$$\text{prox}_{\rho g^*}(x) = x - \rho \text{prox}_{\rho^{-1}g}(\rho^{-1}x),$$

we conclude that

$$\begin{aligned} R_{\rho g^*}(x) &= 2\text{prox}_{\rho g^*}(x) - x \\ &= 2(x - \rho \text{prox}_{\rho^{-1}g}(\rho^{-1}x)) - x \\ &= -\rho \left( 2(\text{prox}_{\rho^{-1}g}(\rho^{-1}x)) - (\rho^{-1}x) \right) \\ &= -\rho R_{\rho^{-1}g}(\rho^{-1}x) \end{aligned}$$

and

$$\begin{aligned} R_{\rho(g^* \circ -\text{Id})}(x) &= 2\text{prox}_{\rho(g^* \circ -\text{Id})}(x) - x \\ &= -2\text{prox}_{\rho g^*}(-x) - x \\ &= -2(-x - \rho \text{prox}_{\rho^{-1}g}(-\rho^{-1}x)) - x \\ &= 2\rho \text{prox}_{\rho^{-1}g}(-\rho^{-1}x) + x \\ &= \rho(2\text{prox}_{\rho^{-1}g}(-\rho^{-1}x) - (-\rho^{-1}x)) \\ &= \rho R_{\rho^{-1}g}(-\rho^{-1}x). \end{aligned}$$

To show the third claim, we first derive an expression for  $r_{\rho(g^* \circ -\text{Id})}^*$ . We have

$$r_{\rho(g^* \circ -\text{Id})}^*(y) = (\rho(g^* \circ -\text{Id}) + \frac{1}{2}\|\cdot\|^2)^*(y)$$

$$\begin{aligned}
 &= \sup_z \{ \langle y, z \rangle - \rho \sup_x \{ \langle z, x \rangle - g(-x) \} - \frac{1}{2} \|z\|^2 \} \\
 &= \sup_z \{ \langle y, z \rangle + \rho \inf_x \{ \langle z, -x \rangle + g(-x) \} - \frac{1}{2} \|z\|^2 \} \\
 &= \sup_z \{ \langle y, z \rangle + \rho \inf_v \{ \langle z, v \rangle + g(v) \} - \frac{1}{2} \|z\|^2 \} \\
 &= \sup_z \inf_v \{ \langle y, z \rangle + \rho \langle z, v \rangle + \rho g(v) - \frac{1}{2} \|z\|^2 \} \\
 &= \inf_v \sup_z \{ \langle y + \rho v, z \rangle + \rho g(v) - \frac{1}{2} \|z\|^2 \} \\
 &= \inf_v \{ \frac{1}{2} \|y + \rho v\|^2 + \rho g(v) \} \\
 &= \inf_v \{ \langle y, \rho v \rangle + \frac{1}{2} \|\rho v\|^2 + \rho g(v) \} + \frac{1}{2} \|y\|^2 \\
 &= - \sup_v \{ \langle -y, \rho v \rangle - \frac{1}{2} \|\rho v\|^2 - \rho g(v) \} + \frac{1}{2} \|y\|^2 \\
 &= -\rho^2 \sup_v \{ \langle -\rho^{-1}y, v \rangle - \frac{1}{2} \|v\|^2 - \rho^{-1}g(v) \} + \frac{1}{2} \|y\|^2 \\
 &= -\rho^2 r_{\rho^{-1}g}^* (-\rho^{-1}y) + \frac{1}{2} \|y\|^2,
 \end{aligned}$$

where the sup-inf swap is valid by the minimax theorem in [39], since we can construct a compact set for the  $z$  variable due to strong convexity of  $\| \cdot \|^2$ . This implies that

$$\begin{aligned}
 p_{\rho(g^* \circ -\text{Id})}(y) &= 2r_{\rho(g^* \circ -\text{Id})}^*(y) - \frac{1}{2} \|y\|^2 \\
 &= -2\rho^2 r_{\rho^{-1}g}^* (-\rho^{-1}y) + \frac{1}{2} \|y\|^2 \\
 &= -\rho^2 (2r_{\rho^{-1}g}^* (-\rho^{-1}y) - \frac{1}{2} \|-\rho^{-1}y\|^2) \\
 &= -\rho^2 p_{\rho^{-1}g} (-\rho^{-1}y).
 \end{aligned}$$

This concludes the proof.

### Appendix D: Technical Lemmas

**Lemma D.1** *Assume that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable and that  $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  are linear operators. Then,*

$$-\frac{1}{2} \langle M(x - y), x - y \rangle \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{1}{2} \langle L(x - y), x - y \rangle \tag{29}$$

if and only if

$$-\langle M(x - y), x - y \rangle \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \langle L(x - y), x - y \rangle. \tag{30}$$

*Proof* Adding two copies of (29) with  $x$  and  $y$  interchanged gives

$$-\langle M(x - y), x - y \rangle \leq \langle \nabla f(x) - f(y), x - y \rangle \leq \langle L(x - y), x - y \rangle. \tag{31}$$

This shows that (29) implies (30). To show the other direction, we use integration. Let  $h(\tau) = f(x + \tau(y - x))$ , then

$$\nabla h(\tau) = \langle y - x, \nabla f(x + \tau(y - x)) \rangle.$$

Since  $f(y) = h(1)$  and  $f(x) = h(0)$ , we get

$$f(y) - f(x) = h(1) - h(0) = \int_0^1 \nabla h(\tau) d\tau = \int_0^1 \langle y - x, \nabla f(x + \tau(y - x)) \rangle d\tau.$$

Therefore

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x \rangle &= \int_0^1 \langle \nabla f(x + \tau(y - x)), y - x \rangle d\tau \\ &\quad - \langle \nabla f(x), y - x \rangle \\ &= \int_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle d\tau \\ &= \int_0^1 \tau^{-1} \langle \nabla f(x + \tau(y - x)) \\ &\quad - \nabla f(x), \tau(y - x) \rangle d\tau \\ &= \int_0^1 \tau^{-1} \langle \nabla f(x + \tau(y - x)) \\ &\quad - \nabla f(x), (x + \tau(y - x)) - x \rangle d\tau. \end{aligned}$$

Using the upper bound in (30), we get

$$\begin{aligned} &\int_0^1 \tau^{-1} \langle \nabla f(x + \tau(y - x)) - \nabla f(x), (x + \tau(y - x)) - x \rangle d\tau \\ &\leq \int_0^1 \tau^{-1} \langle L\tau(x - y), \tau(x - y) \rangle d\tau \\ &= \langle L(x - y), x - y \rangle \int_0^1 \tau d\tau \\ &= \frac{1}{2} \langle L(x - y), x - y \rangle. \end{aligned}$$

Similarly, using the lower bound in (30), we get

$$\begin{aligned} &\int_0^1 \tau^{-1} \langle \nabla f(x + \tau(y - x)) - \nabla f(x), (x + \tau(y - x)) - x \rangle d\tau \\ &\geq - \int_0^1 \tau^{-1} \langle M\tau(x - y), \tau(x - y) \rangle d\tau \end{aligned}$$

$$\begin{aligned} &= -\langle M(x - y), x - y \rangle \int_0^1 \tau \, d\tau \\ &= -\frac{1}{2} \langle M(x - y), x - y \rangle. \end{aligned}$$

This concludes the proof. □

**Lemma D.2** *Assume that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable and that  $L$  is positive definite. Then, that  $f$  is  $L$ -smooth, i.e., that  $f$  satisfies*

$$|f(x) - f(y) - \langle \nabla f(y), x - y \rangle| \leq \frac{\beta}{2} \|x - y\|_L^2 \tag{32}$$

for all  $x, y \in \mathbb{R}^n$ , is equivalent to that  $\nabla f$  is  $\beta$ -Lipschitz continuous w.r.t.  $\|\cdot\|_L$ , i.e., that

$$\|\nabla f(x) - \nabla f(y)\|_{L^{-1}} \leq \beta \|x - y\|_L \tag{33}$$

holds for all  $x, y \in \mathbb{R}^n$ .

*Proof* We start by proving the result in the induced norm  $\|\cdot\|$ , i.e., with  $L = \text{Id}$ . For this, we introduce the functions  $h := \frac{1}{\beta} f$  and  $r := \frac{1}{2}(h + \frac{1}{2}\|\cdot\|^2)$ .

Since  $L = \text{Id}$ , the condition (33) is  $\beta$ -Lipschitz continuity of  $\nabla f$  (w.r.t.  $\|\cdot\|$ ). This is equivalent to that  $\nabla h = \frac{1}{\beta} \nabla f$  is nonexpansive, which by [28, Proposition 4.2] is equivalent to that  $\frac{1}{2}(\nabla h + \text{Id}) = \nabla(\frac{1}{2}(h + \frac{1}{2}\|\cdot\|^2)) = \nabla r$  is firmly nonexpansive (or equivalently 1-cocoercive). This, in turn, is equivalent to (see [30, Theorem 2.1.5] and [28, Definition 4.4]):

$$0 \leq r(x) - r(y) - \langle \nabla r(y), x - y \rangle \leq \frac{1}{2} \|x - y\|^2$$

for all  $x, y \in \mathbb{R}^n$ . Multiplying by 2 and using  $2r = h + \frac{1}{2}\|\cdot\|^2$ , gives

$$\begin{aligned} 0 &\leq h(x) - h(y) - \langle \nabla h(y), x - y \rangle + \frac{1}{2} (\|x\|^2 - \|y\|^2 - 2\langle y, x - y \rangle) \\ &= h(x) - h(y) - \langle \nabla h(y), x - y \rangle + \frac{1}{2} \|x - y\|^2 \leq \|x - y\|^2. \end{aligned}$$

Multiplying by  $\beta$  and using  $f = \beta h$ , we obtain

$$-\frac{\beta}{2} \|x - y\| \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{\beta}{2} \|x - y\|^2.$$

This chain of equivalences show that the conditions are equivalent when  $L = \text{Id}$ .

It remains to show that the scaled version holds. For this, we introduce the function  $g = f \circ L^{-1/2}$ . Letting  $u = L^{-1/2}x$  and  $v = L^{-1/2}y$ , we get  $g(x) = f(u)$ ,  $g(y) = f(v)$ , and  $\nabla g(y) = L^{-1/2} \nabla f(v)$ . Inserting these into the inequality (32) with  $L = \text{Id}$  applied to  $g$  shows (with some simple algebra) that it reduces to the stated inequality (32) in  $f$  and  $L$ . Similarly, the inequality (33) with  $L = \text{Id}$  applied to  $g$  reduces to the stated inequality (33) in  $f$  and  $L$ . This concludes the proof. □

**Lemma D.3** *Suppose that  $P$  is a linear self-adjoint and nonexpansive operator with largest eigenvalue  $\lambda_{\max}(P) = L$  and smallest eigenvalue  $\lambda_{\min}(P) = m$ , satisfying  $-1 \leq m \leq L \leq 1$ , and suppose that  $\delta \in [-1, 1]$  and let  $j$  be the index that minimizes  $|\frac{1}{2\delta} - \lambda_i(P)|$ . The smallest eigenvalue of  $P - \delta P^2$  satisfies the following:*

- (i) if  $\delta \in [0, 1]$ , then  $\lambda_{\min}(P - \delta P^2) = \min(m - \delta m^2, L - \delta L^2)$ .
- (ii) if  $\delta \in [-0.5, 0]$ , then  $\lambda_{\min}(P - \delta P^2) = m - \delta m^2$ .
- (iii) if  $\delta \in [-1, -0.5]$ , then  $\lambda_{\min}(P - \delta P^2) = \lambda_j(P) - \delta \lambda_j(P)^2$ , where  $j = \operatorname{argmin}_i (|\frac{1}{2\delta} - \lambda_i(P)|)$ .

*The largest eigenvalue of  $P + \delta P^2$  satisfies the following:*

- (li) if  $\delta \in [-0.5, 1]$ , then  $\lambda_{\max}(P + \delta P^2) = L + \delta L^2$ .
- (lii) if  $\delta \in [-1, -0.5]$ , then  $\lambda_{\max}(P + \delta P^2) = \lambda_j(P) + \delta \lambda_j(P)^2$ , where  $j = \operatorname{argmin}_i (|\frac{1}{2\delta} + \lambda_i(P)|)$ .

*Proof* The spectral theorem implies that  $\lambda_i(P - \delta P^2) = \lambda_i(P) - \delta \lambda_i(P)^2$ . Therefore, we need to find the eigenvalues  $\lambda_i(P)$  that minimizes the function  $\psi(\lambda) = \lambda - \delta \lambda^2$ , where  $\lambda_i(P) \in [-1, 1]$  for different  $\delta \in [-1, 1]$ .

- (i) For  $\delta \in [0, 1]$ , the function  $\psi$  is concave, and the minimum is found in either of the end points, so  $\lambda_{\min}(P - \delta P^2) = \min(m - \delta m^2, L - \delta L^2)$ .

For  $\delta \in [-1, 0[$  the function  $\psi$  is convex. The unconstrained minimum is at  $\frac{1}{2\delta}$ . The level sets of  $\psi$  are symmetric around  $\frac{1}{2\delta}$ . Therefore, the constrained minimum is the eigenvalue  $\lambda_i(P)$  closest to  $\frac{1}{2\delta}$ :

- (i) For  $\delta \in [-0.5, 0[$ ,  $\lambda_{\min}(P) = m$
- (ii) For  $\delta \in [-1, -0.5]$ ,  $\lambda_{\min}(P) = \lambda_j(P)$ .

To show the largest eigenvalues of  $P + \delta P^2$ , we proceed analogously to the above. Details are omitted.  $\square$

## References

- Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization* **53**(5–6), 475–504 (2004)
- Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. *Trans. Am. Math. Soc.* **82**, 421–439 (1956)
- Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**(6), 964–979 (1979)
- Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* **2**(1), 17–40 (1976)
- Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Math. Model. Numer. Anal.* **9**, 41–76 (1975)
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
- Chambolle, A., Pock, T.: A first-order primal–dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.* **40**(1), 120–145 (2011)



8. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications (2015). [arXiv:1504.01032](https://arxiv.org/abs/1504.01032)
9. Gubin, L.G., Polyak, B.T., Raik, E.V.: The method of projections for finding the common point of convex sets. *USSR Comput. Math. Math. Phys.* **7**(6), 1–24 (1967)
10. Agmon, S.: The relaxation method for linear inequalities. *Can. J. Math.* **6**(3), 382–392 (1954)
11. Motzkin, T.S., Shoenberg, I.: The relaxation method for linear inequalities. *Can. J. Math.* **6**(3), 383–404 (1954)
12. Eremin, I.I.: Generalization of the Motzkin–Agmon relaxation method. *Usp. mat. Nauk* **20**(2), 183–188 (1965)
13. Bregman, L.M.: Finding the common point of convex sets by the method of successive projection. *Dokl. Akad. Nauk SSSR* **162**(3), 487–490 (1965)
14. von Neumann, J.: *Functional Operators. Volume II. The Geometry of Orthogonal Spaces*, *Annals of Mathematics Studies*. Princeton University Press, Princeton (1950). (Reprint of 1933 lecture notes)
15. Benzi, M.: Preconditioning techniques for large linear systems: a survey. *J. Comput. Phys.* **182**(2), 418–477 (2002)
16. Bramble, J.H., Pasciak, J.E., Vassilev, A.T.: Analysis of the inexact Uzawa algorithm for saddle point problems. *SIAM J. Numer. Anal.* **34**(3), 1072–1092 (1997)
17. Hu, Q., Zou, J.: Nonlinear inexact Uzawa algorithms for linear and nonlinear saddle-point problems. *SIAM J. Optim.* **16**(3), 798–825 (2006)
18. Ghadimi, E., Teixeira, A., Shames, I., Johansson, M.: Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems. *IEEE Trans. Autom. Control* **60**(3), 644–658 (2015)
19. Giselsson, P., Boyd, S.: Metric selection in fast dual forward–backward splitting. *Automatica* **62**, 1–10 (2015)
20. Giselsson, P., Boyd, S.: Linear convergence and metric selection for Douglas–Rachford splitting and ADMM. *IEEE Trans. Autom. Control* **62**(2), 532–544 (2017)
21. Giselsson, P.: Tight global linear convergence rate bounds for Douglas–Rachford splitting. *J. Fixed Point Theory Appl.* (2017). <https://doi.org/10.1007/s11784-017-0417-1>
22. Patrinos, P., Stella, L., Bemporad, A.: Forward–backward truncated Newton methods for convex composite optimization. (2014). [arXiv:1402.6655](https://arxiv.org/abs/1402.6655)
23. Stella, L., Themelis, A., Patrinos, P.: Forward–backward quasi-Newton methods for nonsmooth optimization problems. *Comp. Opt. and Appl.* **67**(3), 443–487 (2017)
24. Patrinos, P., Stella, L., Bemporad, A.: Douglas–Rachford splitting: complexity estimates and accelerated variants. In: *Proceedings of the 53rd IEEE Conference on Decision and Control*, pp. 4234–4239. Los Angeles, CA (2014)
25. Themelis, A., Stella, L., Patrinos, P.: Forward–backward envelope for the sum of two nonconvex functions: further properties and nonmonotone line-search algorithms. (2016). [arXiv:1606.06256](https://arxiv.org/abs/1606.06256)
26. Themelis, A., Stella, L., Patrinos, P.: Douglas–Rachford splitting and ADMM for nonconvex optimization: new convergence results and accelerated versions. (2017). [arXiv:1709.05747](https://arxiv.org/abs/1709.05747)
27. Pejčić, I., Jones, C.N.: Accelerated ADMM based on accelerated Douglas–Rachford splitting. In: *2016 European Control Conference (ECC)*, pp. 1952–1957 (2016)
28. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York (2011)
29. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer, Berlin (1998)
30. Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*, 1st edn. Springer, Dordrecht (2003)
31. Nocedal, J., Wright, S.: *Numerical Optimization*. Springer series in operations research and financial engineering, 2nd edn. Springer, New York (2006)
32. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France* **93**, 273–299 (1965)
33. Rockafellar, R.T.: *Convex Analysis*, vol. 28. Princeton University Press, Princeton (1970)
34. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*. North-Holland, Amsterdam (1983)
35. Eckstein, J.: *Splitting methods for monotone operators with applications to parallel optimization*. Ph.D. thesis, MIT (1989)
36. Giselsson, P., Fält, M., Boyd, S.: Line search for averaged operator iteration. (2016). [arXiv:1603.06772](https://arxiv.org/abs/1603.06772)

37. Giselsson, P., Fält, M., Boyd, S.: Line search for averaged operator iteration. In: Proceedings of the 55th Conference on Decision and Control. Las Vegas, USA (2016)
38. Clarke, F.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
39. Sion, M.: On general minimax theorems. *Pac. J. Math.* **8**(1), 171–176 (1958)