

On Carlier's Inequality

Heinz H. Bauschke, Shambhavi Singh, Xianfu Wang

Department of Mathematics, University of British Columbia, Kelowna, Canada
heinz.bauschke@ubc.ca, sambha@student.ubc.ca, shawn.wang@ubc.ca

Dedicated to Roger J-B Wets on the occasion of his 85th birthday.

Received: June 11, 2022

Accepted: February 25, 2023

The Fenchel-Young inequality is fundamental in Convex Analysis and Optimization. It states that the difference between certain function values of two vectors and their inner product is nonnegative. Recently, Carlier introduced a very nice sharpening of this inequality, providing a lower bound that depends on a positive parameter. In this note, we expand on Carlier's inequality in three ways. First, a duality statement is provided. Secondly, we discuss asymptotic behaviour as the underlying parameter approaches zero or infinity. Thirdly, relying on cyclic monotonicity and associated Fitzpatrick functions, we present a lower bound that features an infinite series of squares of norms. Several examples illustrate our results.

Keywords: Carlier's inequality, cyclic monotonicity, Fenchel conjugate, Fenchel-Young inequality, Fitzpatrick function, maximally monotone operator, proximal mapping, resolvent.

2020 Mathematics Subject Classification: Primary 26B25, 47H05; secondary 26D07, 90C25.

1. Introduction

Throughout the paper, we assume that X is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. We also assume throughout that

$$f: X \rightarrow]-\infty, +\infty] \text{ is convex, lower semicontinuous, and proper,} \quad (1)$$

and that $A: X \rightrightarrows X$ is a maximally monotone operator on X . (2)

The *Fenchel conjugate* f^* of f is defined by $f^*(x^*) = \sup_{x \in X} (\langle x, x^* \rangle - f(x))$. The classical Fenchel-Young inequality states that for x and x^* in X , we have

$$G(x, x^*) := G_f(x, x^*) := f(x) + f^*(x^*) - \langle x, x^* \rangle \geq 0, \quad (3)$$

and we have the well known equality characterization

$$G(x, x^*) = 0 \Leftrightarrow x^* \in \partial f(x). \quad (4)$$

(We assume the reader has some basic knowledge of convex analysis and monotone operator theory as can be found, e.g., in [3], [13], [14], and [15].) In [8], Carlier proved recently the following stunningly beautiful sharpening of (3):

$$f(x) + f^*(x^*) - \langle x, x^* \rangle \geq \frac{\|x - \text{Prox}_{\gamma f}(x + \gamma x^*)\|^2}{\gamma}, \quad (5)$$

where $\gamma > 0$ and $\text{Prox}_{\gamma f}$ is the proximal mapping of γf . He also discusses applications and connections to optimal transport, the Brøndsted-Rockafellar theorem, and tilted duality.

The aim of this paper is to expand on Carlier's work in three ways: (1) duality (see Theorem 4.5), (2) asymptotic behaviour (see Theorem 5.1 and Corollary 5.2), and (3) cyclic monotonicity (Corollary 7.4).

The remainder of this paper is organized as follows. In Section 2, inequalities are provided based on the Fitzpatrick function. Useful identities involving the Minty parametrization are presented in Section 3. Carlier's inequality and a new duality result are given in Section 4. In Section 5, we discuss the behaviour of the right side of (5) when $\gamma \rightarrow 0^+$ and $\gamma \rightarrow +\infty$. Various examples are presented in Section 6 to illustrate our results. In Section 7, we obtain sharpenings when the underlying operator A is cyclically monotone of an order bigger than 2.

The notation employed in this paper is fairly standard and follows largely [3].

2. The Fitzpatrick function

In this section, we start the approach to Carlier's result. Several of the proofs are implicit in Carlier's work; however, we include for completeness and the reader's convenience. Recall that the *Fitzpatrick function* for the operator A (see (2)) at $(x, x^*) \in X \times X$ is given by

$$F_A(x, x^*) := \langle x, x^* \rangle - \inf_{(a, a^*) \in \text{gra } A} \langle x - a, x^* - a^* \rangle \quad (6a)$$

$$= \sup_{(a, a^*) \in \text{gra } A} (\langle x, a^* \rangle + \langle a, x^* \rangle - \langle a, a^* \rangle); \quad (6b)$$

see [12] for the original paper and also [16] for various extensions, applications, and further references. It is known (see [12]) that

$$F_A(x, x^*) \geq \langle x, x^* \rangle, \quad \text{with equality if and only if } x^* \in Ax. \quad (7)$$

The next result will be useful later.

Proposition 2.1. *Given x, y, x^*, y^* in X , we have*

$$F_A(x, y^*) + F_A(y, x^*) - \langle x, x^* \rangle - \langle y, y^* \rangle \geq \langle y - x, x^* - y^* \rangle, \quad (8)$$

and equality holds if and only if $y^ \in Ax$ and $x^* \in Ay$.*

Proof. By (7), we have

$$F_A(x, y^*) \geq \langle x, y^* \rangle \quad \text{and} \quad F_A(y, x^*) \geq \langle y, x^* \rangle; \quad (9)$$

moreover, equality holds for both inequalities if and only if $y^* \in Ax$ and $x^* \in Ay$. Adding the two inequalities in (9), followed by subtracting $\langle x, x^* \rangle + \langle y, y^* \rangle$ from both sides, gives, as claimed,

$$\begin{aligned} & F_A(x, y^*) + F_A(y, x^*) - \langle x, x^* \rangle - \langle y, y^* \rangle \\ & \geq \langle x, y^* \rangle + \langle y, x^* \rangle - \langle x, x^* \rangle - \langle y, y^* \rangle = \langle y - x, x^* - y^* \rangle. \quad \square \end{aligned}$$

Fact 2.2. (Fitzpatrick) *Define $(f \oplus f^*)(x, x^*) := f(x) + f^*(x^*)$. Then we have $f \oplus f^* \geq F_{\partial f}$.*

Proof. This is contained in [12]; see also the discussion in [4, Section 2]. For completeness, we include the short proof here. Let $(a, a^*) \in \text{gra } \partial f$. Then

$$\langle a, x^* \rangle + \langle x, a^* \rangle - \langle a, a^* \rangle = (\langle a, x^* \rangle - f(a)) + (\langle x, a^* \rangle - f^*(a^*)) \quad (10a)$$

$$\leq f^*(x^*) + f(x). \quad (10b)$$

The result follows by taking the supremum over $(a, a^*) \in \text{gra } \partial f$. $\square$

Corollary 2.3. (Carlier) *Given x, y, x^*, y^* in X , we have*

$$G_f(x, x^*) + G_f(y, y^*) \geq \langle y - x, x^* - y^* \rangle \quad (11)$$

and

$$G_f(x, x^*) + G_f(y, y^*) = \langle y - x, x^* - y^* \rangle \Leftrightarrow [y^* \in \partial f(x) \text{ and } x^* \in \partial f(y)]. \quad (12)$$

Proof. (See also [8, Section 1].) Using Fact 2.2 and Proposition 2.1, we always have

$$G_f(x, x^*) + G_f(y, y^*) = (f(x) + f^*(x^*) - \langle x, x^* \rangle) + (f(y) + f^*(y^*) - \langle y, y^* \rangle) \quad (13a)$$

$$= (f(x) + f^*(y^*)) + (f(y) + f^*(x^*)) - \langle x, x^* \rangle - \langle y, y^* \rangle \quad (13b)$$

$$\geq F_{\partial f}(x, y^*) + F_{\partial f}(y, x^*) - \langle x, x^* \rangle - \langle y, y^* \rangle \quad (13c)$$

$$\geq \langle y - x, x^* - y^* \rangle, \quad (13d)$$

which is (11). We now turn to the proof of (12).

“ $\Rightarrow$ ”: In this case, we have equality in (13d). In turn, the equality characterization in Proposition 2.1 yields $y^* \in \partial f(x)$ and $x^* \in \partial f(y)$.

“ $\Leftarrow$ ”: In this case, $f(x) + f^*(y^*) = \langle x, y^* \rangle$ and $f(y) + f^*(x^*) = \langle y, x^* \rangle$. It follows

$$(f(x) + f^*(y^*)) + (f(y) + f^*(x^*)) - \langle x, x^* \rangle - \langle y, y^* \rangle \quad (14)$$

$$= \langle x, y^* \rangle + \langle y, x^* \rangle - \langle x, x^* \rangle - \langle y, y^* \rangle = \langle y - x, x^* - y^* \rangle. \quad (15)$$

Hence the chain of inequalities in (13) is actually a chain of equalities and we are done. $\square$

3. Minty parametrization

In this section, we employ the Minty parametrization and derive results that will be useful later. Recalling the standing assumption (2), and given points x, x^* in X and $\gamma > 0$, we have the well known equivalences (see, e.g., [3, Chapter 23]):

$$x^* \in Ax \Leftrightarrow x + \gamma x^* \in x + \gamma Ax \Leftrightarrow x = J_{\gamma A}(x + \gamma x^*), \quad (16)$$

where $J_{\gamma A} = (\text{Id} + \gamma A)^{-1}$ is the resolvent of γA . The Yosida approximation is denoted by ${}^\gamma A x := (x - J_{\gamma A} x) / \gamma$.

Lemma 3.1. *Let x, x^* be in X , and let $\gamma > 0$. Set*

$$a := J_{\gamma A}(x + \gamma x^*) \quad \text{and} \quad a^* := \frac{x + \gamma x^* - J_{\gamma A}(x + \gamma x^*)}{\gamma}. \quad (17)$$

$$\text{Then} \quad x + \gamma x^* = a + \gamma a^*, \quad (18)$$

$$(a, a^*) \in \text{gra } A, \quad (19)$$

$$\text{and} \quad \langle a - x, x^* - a^* \rangle = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma}. \quad (20)$$

$$\text{Moreover,} \quad \|(x, x^*) - (a, a^*)\|^2 = \left(1 + \frac{1}{\gamma^2}\right) \|x - J_{\gamma A}(x + \gamma x^*)\|^2 \quad (21)$$

$$\text{and} \quad \|(x - a) - (x^* - a^*)\|^2 = \left(1 + \frac{1}{\gamma}\right)^2 \|x - J_{\gamma A}(x + \gamma x^*)\|^2. \quad (22)$$

Finally, we have the equivalences (which may or may not hold)

$$x^* \in Ax \Leftrightarrow a = x \Leftrightarrow a^* = x^* \Leftrightarrow \langle a - x, x^* - a^* \rangle = 0. \quad (23)$$

Proof. Clearly, (17) implies (18). By applying the classical Minty parametrization to γA , we have $(a, \gamma a^*) \in \text{gra}(\gamma A)$ and thus $(a, a^*) \in \text{gra } A$, i.e., (19) holds. Note that

$$x - a = x - J_{\gamma A}(x + \gamma x^*) \quad (24)$$

$$\text{and} \quad x^* - a^* = \frac{\gamma x^* - \gamma a^*}{\gamma} = \frac{\gamma x^* - (x + \gamma x^* - J_{\gamma A}(x + \gamma x^*))}{\gamma} \quad (25a)$$

$$= -\frac{1}{\gamma}(x - J_{\gamma A}(x + \gamma x^*)). \quad (25b)$$

Combining (24) and (25), we deduce

$$-\langle x - a, x^* - a^* \rangle = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma}, \quad (26)$$

which is (20). Next, using (24) and (25) again, we obtain

$$\|(x, x^*) - (a, a^*)\|^2 = \|x - a\|^2 + \|x^* - a^*\|^2 \quad (27a)$$

$$= \|x - J_{\gamma A}(x + \gamma x^*)\|^2 + \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma^2}, \quad (27b)$$

which yields (21). Moreover, using (24), (25), and (26), we deduce that

$$\|(x - a) - (x^* - a^*)\|^2 = \|x - a\|^2 + \|x^* - a^*\|^2 - 2\langle x - a, x^* - a^* \rangle \quad (28a)$$

$$= \left(1 + \frac{1}{\gamma^2} + \frac{2}{\gamma}\right) \|x - J_{\gamma A}(x + \gamma x^*)\|^2, \quad (28b)$$

which yields (22). We now turn to (23). We rewrite (27) as

$$\|x - a\|^2 + \|x^* - a^*\|^2 = \left(1 + \frac{1}{\gamma^2}\right) \|x - a\|^2.$$

Hence, invoking also (26), we obtain

$$\gamma \|x^* - a^*\|^2 = \frac{1}{\gamma} \|x - a\|^2 = -\langle x - a, x^* - a^* \rangle,$$

which yields the equivalences $x = a \Leftrightarrow x^* = a^* \Leftrightarrow \langle x - a, x^* - a^* \rangle = 0$. If $x = a$, then $(x, x^*) = (a, a^*) \in \text{gra } A$ by (19). And if $(x, x^*) \in \text{gra } A$, then (16) yields $x = a$. All this proves (23). $\square$

We now record a duality result.

Lemma 3.2. *Let x, x^* be in X , and let $\gamma > 0$. Then*

$$x^* - J_{\gamma^{-1}A^{-1}}(x^* + \gamma^{-1}x) = -\gamma^{-1}(x - J_{\gamma A}(x + \gamma x^*)); \quad (29)$$

$$\text{consequently, } \frac{\|x^* - J_{\gamma^{-1}A^{-1}}(x^* + \gamma^{-1}x)\|^2}{\gamma^{-1}} = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma}. \quad (30)$$

Proof. Using [3, Proposition 23.20], we obtain

$$\begin{aligned} x^* - J_{\gamma^{-1}A^{-1}}(x^* + \gamma^{-1}x) &= x^* - (\text{Id} - \gamma^{-1}J_{\gamma A} \circ \gamma \text{Id})(x^* + \gamma^{-1}x) \\ &= x^* - (x^* + \gamma^{-1}x) + \gamma^{-1}J_{\gamma A}(x + \gamma x^*) = -\gamma^{-1}(x - J_{\gamma A}(x + \gamma x^*)), \end{aligned}$$

which is (29) and from which (30) follows. $\square$

4. The Carlier bound and duality

This section contains a review of Carlier's inequality and a new duality result.

Definition 4.1. (Carlier bound) Recall (2), let x and x^* be in X , and let $\gamma > 0$. We define the associated *Carlier bound* by

$$C(x, x^*) := C_{A, \gamma}(x, x^*) := \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma}. \quad (31)$$

$$\text{If } A = \partial f, \text{ then: } C(x, x^*) = C_{\partial f, \gamma}(x, x^*) = \frac{\|x - \text{Prox}_{\gamma f}(x + \gamma x^*)\|^2}{\gamma}. \quad (32)$$

Theorem 4.2. (Carlier) Recall (2), and let x, x^* be in X . Then

$$(\forall \gamma > 0) \quad F_A(x, x^*) - \langle x, x^* \rangle \geq C_{A, \gamma}(x, x^*).$$

Proof. (See also [8, Section 2].) Let $\gamma > 0$ and set

$$a := J_{\gamma A}(x + \gamma x^*) \quad \text{and} \quad a^* := \frac{x + \gamma x^* - J_{\gamma A}(x + \gamma x^*)}{\gamma}.$$

By (6) and (20), we have, as claimed:

$$\begin{aligned} F_A(x, x^*) - \langle x, x^* \rangle &= - \inf_{(b, b^*) \in \text{gra } A} \langle x - b, x^* - b^* \rangle \geq \langle a - x, x^* - a^* \rangle \\ &= \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} = C_{A, \gamma}(x, x^*). \end{aligned} \quad \square$$

Theorem 4.3. (Carlier) Let x and x^* be in X , and let $\gamma > 0$. Then

$$G_f(x, x^*) \geq F_{\partial f}(x, x^*) - \langle x, x^* \rangle \geq C_{\partial f, \gamma}(x, x^*). \quad (33)$$

Moreover, we have the characterization

$$G_f(x, x^*) = C_{\partial f, \gamma}(x, x^*) \Leftrightarrow \quad (34)$$

$$[\text{Prox}_{\gamma f}(x + \gamma x^*) \in \partial f^*(x^*) \quad \text{and} \quad x + \gamma x^* - \text{Prox}_{\gamma f}(x + \gamma x^*) \in \gamma \partial f(x)].$$

Proof. (See also [8, Section 1].) Write $A = \partial f$ so that $A^{-1} = \partial f^*$. Now set

$$a := J_{\gamma A}(x + \gamma x^*) \quad \text{and} \quad a^* := \frac{x + \gamma x^* - J_{\gamma A}(x + \gamma x^*)}{\gamma}. \quad (35)$$

Then (33) follows from Fact 2.2 and Theorem 4.2. We now derive this differently in order to characterize equality. By (19), we have $(a, a^*) \in \text{gra } A$.

Hence, (4) yields $G_f(a, a^*) = 0$.

Applying now (11) and (20) yields

$$G_f(x, x^*) = G_f(x, x^*) + G_f(a, a^*) \geq \langle a - x, x^* - a^* \rangle = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma}. \quad (36)$$

Moreover, thanks to (12), we have equality characterization

$$G_f(x, x^*) = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} \Leftrightarrow [a^* \in Ax \text{ and } x^* \in Aa],$$

which is precisely (34). $\square$

Remark 4.4. In view of Theorem 4.2, the Carlier bound

$$C_{A,\gamma}(x, x^*) = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} \quad (37)$$

is always less sharp than the Fitzpatrick function bound

$$F_A(x, x^*) - \langle x, x^* \rangle; \quad (38)$$

however, Carlier's lower bound is sharper than the trivial lower bound 0. Computing Fitzpatrick functions is not an easy task (see [4]) – there are many more examples of prox operators available (see [3], [5], and [9]). See also [7] and [6] for recent work on Fitzpatrick functions and related objects. $\square$

Let us state a new duality result, which links the Carlier bound of A to that of A^{-1} :

Theorem 4.5. (Duality) *Recall (2), let x, x^* be in X , and let $\gamma > 0$. Then*

$$C_{A,\gamma}(x, x^*) = C_{A^{-1},\gamma^{-1}}(x^*, x). \quad (39)$$

Proof. Combine (31) with (30). $\square$

We conclude this section by outlining another possible area where Carlier's inequality may be useful – Bregman distances!

Remark 4.6. (Bregman distance) Recall (1) and that the Bregman distance between $x \in X$ and $y \in \text{int dom } f$ is defined by

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle. \quad (40)$$

$$\text{Note that} \quad D_f(x, y) = G_f(x, \nabla f(y)) \geq C_{\partial f, \gamma}(x, \nabla f(y)) \quad (41)$$

by Theorem 4.3. The Bregman distance plays a role, e.g., when analyzing the proximal gradient method (PGM). If $(y_n)_{n \in \mathbb{N}}$ is the sequence generated by the PGM and x is a solution, then $\sum_{n \in \mathbb{N}} D_f(x, y_n) < \infty$ (see, e.g., the proof of [5, Theorem 10.21]).

It follows that $D_f(x, y_n) \rightarrow 0$ and also we learn from (41) that

$$\sum_{n \in \mathbb{N}} C_{\partial f, \gamma}(x, \nabla f(y_n)) = \sum_{n \in \mathbb{N}} \frac{\|x - \text{Prox}_{\gamma f}(x + \gamma \nabla f(y_n))\|^2}{\gamma} < +\infty. \quad (42)$$

This is a prototypical appearance of Carlier's inequality in the context of the analysis of algorithms.

5. Asymptotic behaviour

Let us now analyze the behaviour of Carlier's bound

$$C_{A, \gamma}(x, x^*) = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} \quad (43)$$

when $\gamma \rightarrow 0^+$. Because of Theorem 4.5, we also obtain information about the behaviour when $\gamma \rightarrow +\infty$.

Theorem 5.1. *Recall (2), let x, x^* be in X , and let $\gamma > 0$. Then the following hold:*

- (i) *If $x \notin \overline{\text{dom } A}$, then $\lim_{\gamma \rightarrow 0^+} C_{A, \gamma}(x, x^*) = +\infty$.*
- (ii) *If $x \in \text{dom } A$, then $\lim_{\gamma \rightarrow 0^+} C_{A, \gamma}(x, x^*) = 0$.*

Proof. (i) Suppose that $x \notin \overline{\text{dom } A}$. Set $\delta := d_{\overline{\text{dom } A}}(x) > 0$. Because we have $\text{ran } J_{\gamma A} = \text{dom}(\gamma A) = \text{dom } A$, we estimate

$$\frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} \geq \frac{\delta^2}{\gamma}. \quad (44)$$

This yields the conclusion.

- (ii) Suppose that $x \in \text{dom } A$. Recall that the Yosida approximation is given by ${}^\gamma A x = (x - J_{\gamma A} x)/\gamma$. Because resolvents are nonexpansive, we have

$$\|J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*)\| \leq \gamma \|x^*\|. \quad (45)$$

Clearly,

$$\begin{aligned} & \|x - J_{\gamma A}(x + \gamma x^*)\|^2 \\ &= \|x - J_{\gamma A} x\|^2 + \|J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*)\|^2 + 2 \langle x - J_{\gamma A} x, J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*) \rangle \end{aligned}$$

and this implies

$$\begin{aligned} & \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} \\ &= \gamma \|{}^\gamma A x\|^2 + \frac{\|J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} + 2 \langle {}^\gamma A x, J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*) \rangle. \end{aligned} \quad (46)$$

Because $x \in \text{dom } A$, we learn from [3, Corollary 23.46(i)] that

$$\lim_{\gamma \rightarrow 0^+} {}^\gamma A x = {}^0 A x := P_{Ax}(0). \quad (47)$$

Consider the three summands on the right side of (46). It suffices to show that each one of them goes to 0 as $\gamma \rightarrow 0^+$. First, ${}^\gamma A x \rightarrow P_{Ax}(0)$ and thus

$$\gamma \|{}^\gamma A x\|^2 \rightarrow 0 \|P_{Ax}(0)\|^2 = 0.$$

Second, (45) yields $0 \leq (1/\gamma) \|J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*)\|^2 \leq \gamma \|x^*\|^2 \rightarrow 0$ and therefore $(1/\gamma) \|J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*)\|^2 \rightarrow 0$. Thirdly, (45) shows that $J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*) \rightarrow 0$. Combined with (47), we deduce that $2 \langle {}^\gamma A x, J_{\gamma A} x - J_{\gamma A}(x + \gamma x^*) \rangle \rightarrow 0$. $\square$

Corollary 5.2. Recall (2), let x, x^* be in X , and let $\gamma > 0$. Then the following hold:

- (i) If $x^* \notin \overline{\text{ran}} A$, then $\lim_{\gamma \rightarrow +\infty} C_{A,\gamma}(x, x^*) = +\infty$.
- (ii) If $x^* \in \text{ran } A$, then $\lim_{\gamma \rightarrow +\infty} C_{A,\gamma}(x, x^*) = 0$.

Proof. Theorem 4.5 yields $C_{A,\gamma}(x, x^*) = C_{A^{-1},\gamma^{-1}}(x^*, x)$. The result is now clear from Theorem 5.1 (applied to A^{-1}) because $\text{ran } A = \text{dom } A^{-1}$. $\square$

Corollary 5.3. We have $\text{dom} \sup_{\gamma > 0} C_{A,\gamma} \subseteq \overline{\text{dom } A} \times \overline{\text{ran } A}$. (48)

Proof. Combine Theorem 5.1 with Corollary 5.2. $\square$

6. Examples

In this section, we collect several examples to illustrate our results.

Example 6.1. (Indicator of a subspace) Suppose that $A = N_U$, where U is a closed linear subspace of X . By [4, Example 3.1], the Fitzpatrick bound at $(x, x^*) \in X \times X$ is

$$F_{N_U}(x, x^*) - \langle x, x^* \rangle = \iota_U(x) + \iota_{U^\perp}(x^*) - \langle x, x^* \rangle = (\iota_U \oplus \iota_{U^\perp})(x, x^*). \quad (49)$$

Let $\gamma > 0$. Because $J_{\gamma A} = P_U$ is linear, we compute Carlier's bound via

$$C_{N_U,\gamma}(x, x^*) = \frac{\|x - P_U(x + \gamma x^*)\|^2}{\gamma} = \frac{\|P_{U^\perp}x\|^2 + \gamma^2 \|P_U x^*\|^2}{\gamma} \quad (50a)$$

$$= \frac{1}{\gamma} \|P_{U^\perp}x\|^2 + \gamma \|P_U x^*\|^2. \quad (50b)$$

Set $\mathbb{R}_{++} := \{\xi \in \mathbb{R} \mid \xi > 0\}$. Note that $X \times X \times \mathbb{R}_{++} \rightarrow \mathbb{R}: (x, x^*, \gamma) \mapsto C_{N_U,\gamma}(x, x^*)$ is *not* convex; however, the functions $X \times X \rightarrow \mathbb{R}: (x, x^*) \mapsto C_{N_U,\gamma}(x, x^*)$ and $\mathbb{R}_{++} \rightarrow \mathbb{R}: \gamma \mapsto C_{N_U,\gamma}(x, x^*)$ are convex. Let us discuss further

$$\mathbb{R}_{++} \rightarrow \mathbb{R}: \gamma \mapsto C_{N_U,\gamma}(x, x^*). \quad (51)$$

This function is (i) strictly increasing if $x \in U$ and $x^* \notin U^\perp$; (ii) strictly decreasing if $x \notin U$ and $x^* \in U^\perp$; (iii) first strictly decreasing then strictly increasing if $x \notin U$ and $x^* \notin U^\perp$; (iv) identically equal to 0 if $x \in U$ and $x^* \in U^\perp$. Moreover,

$$\lim_{\gamma \rightarrow 0^+} C_{N_U,\gamma}(x, x^*) = \iota_U(x) \quad \text{and} \quad \lim_{\gamma \rightarrow +\infty} C_{N_U,\gamma}(x, x^*) = \iota_{U^\perp}(x^*). \quad (52)$$

$$\text{It follows that} \quad \sup_{\gamma > 0} C_{N_U,\gamma}(x, x^*) = \iota_U(x) + \iota_{U^\perp}(x^*) \quad (53)$$

coincides with the Fitzpatrick bound in this case.

Example 6.2. (Energy) Suppose that $f = \frac{1}{2} \|\cdot\|^2$ and hence $\nabla f = \text{Id}$. By [4, Example 3.10], the Fitzpatrick bound at $(x, x^*) \in X \times X$ is

$$F_{\text{Id}}(x, x^*) - \langle x, x^* \rangle = \frac{1}{4} \|x + x^*\|^2 - \langle x, x^* \rangle = \frac{1}{4} \|x - x^*\|^2. \quad (54)$$

Let $\gamma > 0$. Then $\text{Prox}_{\gamma f} = (\text{Id} + \gamma \text{Id})^{-1} = (1 + \gamma)^{-1} \text{Id}$ and hence Carlier's bound is

$$C_{\text{Id},\gamma}(x, x^*) = \frac{\|x - \text{Prox}_{\gamma f}(x + \gamma x^*)\|^2}{\gamma} = \frac{\|x - (1 + \gamma)^{-1}(x + \gamma x^*)\|^2}{\gamma} \quad (55a)$$

$$= \frac{\gamma}{(1 + \gamma)^2} \|x - x^*\|^2. \quad (55b)$$

If $x \neq x^*$, then the function $\gamma \mapsto C_{\text{Id},\gamma}(x, x^*)$ is strictly concave on $]0, 2]$, strictly convex on $[2, +\infty[$, and its unique global maximizer is $\gamma = 1$ for which we have $C_{\text{Id},1}(x, x^*) = \frac{1}{4} \|x - x^*\|^2 = F_{\text{Id}}(x, x^*)$.

Example 6.3. (Skew rotator) Suppose that $X = \mathbb{R}^2$. We consider the function $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2: (x_1, x_2) \mapsto (-x_2, x_1)$, the counter-clockwise rotator by $\pi/2$, which is a skew isometry. By [2, Proposition 7.4],

$$F_A(x, x^*) = \iota_{\text{gra } A}(x, x^*) = F_A(x, x^*) - \langle x, x^* \rangle. \quad (56)$$

Let $\gamma > 0$. Then

$$J_{\gamma A} = (1 + \gamma^2)^{-1}(\text{Id} - \gamma A): (x_1, x_2) \mapsto (1 + \gamma^2)^{-1}(x_1 + \gamma x_2, -\gamma x_1 + x_2).$$

Thus, after some algebra, we find that Carlier's bound is

$$C_{A,\gamma}(x, x^*) = \frac{\|x - J_{\gamma A}(x + \gamma x^*)\|^2}{\gamma} = \frac{\gamma}{1 + \gamma^2} \|Ax - x^*\|^2. \quad (57)$$

Therefore, if $Ax \neq x^*$, then $\gamma \mapsto C_{A,\gamma}(x, x^*)$ is strictly concave on $]0, \sqrt{3}]$, strictly convex on $[\sqrt{3}, +\infty[$, and its unique global maximizer is $\gamma = 1$ for which we have $C_{A,1}(x, x^*) = \frac{1}{2} \|Ax - x^*\|^2$.

We now focus on the case when $X = \mathbb{R}$ and thus $A = \partial f$. For ease of notation, we will use (x, y) instead of (x, x^*) as we do elsewhere.

Example 6.4. ((Negative) Burg entropy) Suppose that $f(x) = -\ln(x)$ whenever $x > 0$, and $+\infty$ elsewhere, and let $\gamma > 0$. It is known (see, e.g., [5, Example 6.9], [3, Example 24.40], or [10]) that for $z \in \mathbb{R}$,

$$\text{Prox}_{\gamma f}(z) = \frac{z + \sqrt{z^2 + 4\gamma}}{2}. \quad (58)$$

Hence, for $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} C_{\partial f,\gamma}(x, y) &= \frac{|x - \text{Prox}_{\gamma f}(x + \gamma y)|^2}{\gamma} \\ &= \frac{(x - \frac{1}{2}(x + \gamma y) - \frac{1}{2}\sqrt{(x + \gamma y)^2 + 4\gamma})^2}{\gamma} = \frac{((x - \gamma y) - \sqrt{(x + \gamma y)^2 + 4\gamma})^2}{4\gamma}; \end{aligned}$$

in particular,

$$\begin{aligned} C_{\partial f,\gamma}(0, y) &= \frac{((0 - \gamma y) - \sqrt{(0 + \gamma y)^2 + 4\gamma})^2}{4\gamma} = \frac{(\gamma y + \sqrt{(\gamma y)^2 + 4\gamma})^2}{4\gamma} \\ &= \frac{(\sqrt{\gamma}\sqrt{\gamma y} + \sqrt{\gamma}\sqrt{\gamma y^2 + 4})^2}{(2\sqrt{\gamma})^2} = \left(\frac{\sqrt{\gamma y} + \sqrt{\gamma y^2 + 4}}{2} \right)^2 \rightarrow 1 \quad \text{as } \gamma \rightarrow 0^+. \end{aligned} \quad (59)$$

Combining with Theorem 5.1, we obtain

$$\lim_{\gamma \rightarrow 0^+} C_{\partial f, \gamma}(x, y) = \begin{cases} +\infty, & \text{if } x < 0; \\ 1, & \text{if } x = 0; \\ 0, & \text{if } x > 0 \end{cases} \quad (60)$$

which is convex – but not lower semicontinuous – as a function of x .

Example 6.5. ((Negative) Boltzmann-Shannon entropy) Suppose that f at $x \in \mathbb{R}$ is defined by

$$f(x) = \begin{cases} +\infty, & \text{if } x < 0; \\ 0, & \text{if } x = 0; \\ x \ln(x) - x, & \text{if } x > 0. \end{cases} \quad (61)$$

We start by showing that

$$\text{Prox}_{\gamma f}(x) = \gamma W\left(\frac{1}{\gamma} \exp\left(\frac{x}{\gamma}\right)\right) \quad (62)$$

where $x \in \mathbb{R}$ and where W is the *Lambert W-function* which satisfies $W(z)e^{W(z)} = z$ (see [11, Equation 1.5]). To see that, recall that (see, e.g., [3, Proposition 24.1] and also [9]) the characterization of the proximal mapping

$$p = \text{Prox}_{\gamma f}(x) \Leftrightarrow \gamma \nabla f(p) + p = x \Leftrightarrow \gamma \ln(p) + p = x, \quad (63)$$

where $x \in \mathbb{R}$ and $p > 0$. Hence

$$\frac{p}{\gamma} \exp\left(\frac{p}{\gamma}\right) = \frac{1}{\gamma} \exp\left(\frac{x}{\gamma}\right) \Leftrightarrow p = \gamma W\left(\frac{1}{\gamma} \exp\left(\frac{x}{\gamma}\right)\right) \quad (64)$$

by the very definition of the W function, and this verifies (62). Therefore

$$C_{\partial f, \gamma}(x, y) = \frac{|x - \text{Prox}_{\gamma f}(x + \gamma y)|^2}{\gamma} \quad (65a)$$

$$= \frac{\left(x - \gamma W\left(\frac{1}{\gamma} \exp\left(\frac{(x + \gamma y)}{\gamma}\right)\right)\right)^2}{\gamma}. \quad (65b)$$

In particular,

$$C_{\partial f, \gamma}(0, y) = \gamma \left(W\left(\frac{1}{\gamma} \exp(y)\right)\right)^2 = \frac{\left(W\left(\frac{1}{\gamma} \exp(y)\right)\right)^2}{\frac{1}{\gamma}}. \quad (66)$$

We wish to take now the limit as $\gamma \rightarrow 0^+$. As numerator and denominator tend to $+\infty$, we shall use L'Hospital's rule. Using the fact that

$$W'(z) = \frac{1}{(1 + W(z)) \exp(W(z))}, \quad (67)$$

we obtain

$$\lim_{\gamma \rightarrow 0^+} C_{\partial f, \gamma}(0, y) = \lim_{\gamma \rightarrow 0^+} \frac{2W\left(\frac{1}{\gamma} \exp(y)\right) \exp(y)}{\exp\left(W\left(\frac{1}{\gamma} \exp(y)\right)\right) \left(1 + W\left(\frac{1}{\gamma} \exp(y)\right)\right)}. \quad (68)$$

Changing variables via $u = W\left(\frac{1}{\gamma} \exp(y)\right)$, we finally obtain

$$\lim_{\gamma \rightarrow 0^+} C_{\partial f, \gamma}(0, y) = \lim_{u \rightarrow +\infty} \frac{2u \exp(y)}{\exp(u)(1+u)} = 0. \quad (69)$$

Combining this with Theorem 5.1 gives us

$$\lim_{\gamma \rightarrow 0^+} C_{\partial f, \gamma}(x, y) = \begin{cases} +\infty, & \text{if } x < 0; \\ 0, & \text{if } x \geq 0. \end{cases} \quad (70)$$

Remark 6.6. The formulas (59) and (69) illustrate that the asymptotic behaviour at boundary points does not seem to follow a simple pattern and thus warrants further study.

7. Cyclic monotonicity

In this section, we extend the analysis to n -cyclically monotone operators. Recall that A is n -cyclically monotone, where $n \in \{2, 3, \dots\}$, if

$$\left. \begin{array}{l} (a_1, a_1^*) \in \text{gra } A \\ (a_2, a_2^*) \in \text{gra } A \\ \vdots \\ (a_n, a_n^*) \in \text{gra } A \\ a_{n+1} = a_1 \end{array} \right\} \Rightarrow \sum_{k=1}^n \langle a_{k+1} - a_k, a_k^* \rangle \leq 0. \quad (71)$$

2-cyclic monotonicity is just regular monotonicity. The *Fitzpatrick function of order n* , $F_{A,n}$, evaluated at $(x, x^*) \in X \times X$, is the supremum over $(a_1, a_1^*), \dots, (a_{n-1}, a_{n-1}^*)$ in $\text{gra } A$ of the expression

$$\langle x, x^* \rangle + \langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_1 - x, x^* \rangle + \sum_{k=1}^{n-2} \langle a_{k+1} - a_k, a_k^* \rangle. \quad (72)$$

As a supremum of continuous affine functions, the function $F_{A,n}$ is convex and lower semicontinuous. We also set $F_{A,\infty} = \sup_{n \geq 2} F_{A,n}$.

Fact 7.1. (See [1, Corollary 2.8].) *Suppose that A is maximally n -cyclically monotone. Then $F_{A,n} > \langle \cdot, \cdot \rangle$ outside $\text{gra } A$, while $F_{A,n} = \langle \cdot, \cdot \rangle$ on $\text{gra } A$.*

We have (see [1, Remark 2.10]) the ordering

$$\langle \cdot, \cdot \rangle \leq F_{A,2} \leq F_{A,3} \leq \dots \leq F_{A,n} \rightarrow F_{A,\infty}. \quad (73)$$

Moreover, if f is as in (1), then [1, Theorem 3.5] yields for every $(x, x^*) \in X \times X$

$$F_{\partial f, \infty}(x, x^*) = f(x) + f^*(x^*). \quad (74)$$

Computing $F_{A,n}$ is nontrivial; for some concrete examples, see Section 4 in [1] and also [2].

We shall need the following identity.

Lemma 7.2. *Let $(x, x^*) \in X$, and let $(a_1, a_1^*), \dots, (a_{n-1}, a_{n-1}^*)$ be in $X \times X$. Then*

$$\langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_1 - x, x^* \rangle + \sum_{k=1}^{n-2} \langle a_{k+1} - a_k, a_k^* \rangle \quad (75a)$$

$$= \langle a_1 - x, x^* - a_1^* \rangle + \sum_{k=2}^{n-1} \langle a_k - x, a_{k-1}^* - a_k^* \rangle. \quad (75b)$$

Proof. We prove this by induction on $n \geq 2$. If $n = 2$, then the left side of (75) is

$$\langle x - a_1, a_1^* \rangle + \langle a_1 - x, x^* \rangle = \langle a_1 - x, x^* - a_1^* \rangle, \quad (76)$$

which is also equal to the right side of (75).

Now assume the result is true for some $n \geq 2$. We will show the result is also true for $n + 1$.

Indeed, using the inductive hypothesis in (77d), we have

$$\langle x - a_n, a_n^* \rangle + \langle a_1 - x, x^* \rangle + \sum_{k=1}^{n-1} \langle a_{k+1} - a_k, a_k^* \rangle \quad (77a)$$

$$= \langle x - a_n, a_n^* \rangle - \langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_n - a_{n-1}, a_{n-1}^* \rangle \quad (77b)$$

$$+ \langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_1 - x, x^* \rangle + \sum_{k=1}^{n-2} \langle a_{k+1} - a_k, a_k^* \rangle \quad (77c)$$

$$= \langle x - a_n, a_n^* \rangle - \langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_n - a_{n-1}, a_{n-1}^* \rangle \quad (77d)$$

$$+ \langle a_1 - x, x^* - a_1^* \rangle + \sum_{k=2}^{n-1} \langle a_k - x, a_{k-1}^* - a_k^* \rangle \quad (77e)$$

$$= \langle x - a_n, a_n^* \rangle - \langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_n - a_{n-1}, a_{n-1}^* \rangle \quad (77f)$$

$$- \langle a_n - x, a_{n-1}^* - a_n^* \rangle + \langle a_1 - x, x^* - a_1^* \rangle + \sum_{k=2}^n \langle a_k - x, a_{k-1}^* - a_k^* \rangle \quad (77g)$$

$$= \langle a_1 - x, x^* - a_1^* \rangle + \sum_{k=2}^n \langle a_k - x, a_{k-1}^* - a_k^* \rangle \quad (77h)$$

and we are done. □

Theorem 7.3. *Let $(x, x^*) \in X \times X$.*

If $(a_1, a_1^), \dots, (a_{n-1}, a_{n-1}^*)$ belong to $\text{gra } A$, then*

$$F_{A,n}(x, x^*) - \langle x, x^* \rangle \geq \langle a_1 - x, x^* - a_1^* \rangle + \sum_{k=2}^{n-1} \langle a_k - x, a_{k-1}^* - a_k^* \rangle. \quad (78)$$

Moreover, let $\gamma_1 > 0, \dots, \gamma_{n-1} > 0$ and set

$$a_1 := J_{\gamma_1 A}(x + \gamma_1 x^*), \quad a_1^* := \frac{x + \gamma_1 x^* - a_1}{\gamma_1} \quad (79a)$$

$$a_2 := J_{\gamma_2 A}(x + \gamma_2 a_1^*), \quad a_2^* := \frac{x + \gamma_2 a_1^* - a_2}{\gamma_2} \quad (79b)$$

$$\vdots \quad \quad \quad \vdots \quad (79c)$$

$$a_{n-1} := J_{\gamma_{n-1} A}(x + \gamma_{n-1} a_{n-2}^*), \quad a_{n-1}^* := \frac{x + \gamma_{n-1} a_{n-2}^* - a_{n-1}}{\gamma_{n-1}}. \quad (79d)$$

Then

$$F_{A,n}(x, x^*) - \langle x, x^* \rangle \geq \frac{\|x - J_{\gamma_1 A}(x + \gamma_1 x^*)\|^2}{\gamma_1} + \sum_{k=2}^{n-1} \frac{\|x - J_{\gamma_k A}(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k}. \quad (80)$$

Proof. Indeed, using (72) and (75), we have

$$F_{A,n}(x, x^*) - \langle x, x^* \rangle \geq \langle x - a_{n-1}, a_{n-1}^* \rangle + \langle a_1 - x, x^* \rangle + \sum_{k=1}^{n-2} \langle a_{k+1} - a_k, a_k^* \rangle \quad (81a)$$

$$= \langle a_1 - x, x^* - a_1^* \rangle + \sum_{k=2}^{n-1} \langle a_k - x, a_{k-1}^* - a_k^* \rangle \quad (81b)$$

which is (78).

We now turn towards the “Moreover” part.

By Lemma 3.1, the pairs $(a_1, a_1^*), \dots, (a_{n-1}, a_{n-1}^*)$ lie in $\text{gra } A$. The inequality (80) follows by combining (78) with (20). $\square$

Corollary 7.4. (A series lower bound) *Recall that f satisfies (1), and let x, x^* be in X . Let $(\gamma_n)_{n \geq 1}$ be a sequence in $\mathbb{R}_{++}$. Generate $(a_n)_{n \geq 1}$ and $(a_n^*)_{n \geq 0}$ via*

$$a_0^* := x^*, \quad (\forall n \geq 1) \quad a_n := \text{Prox}_{\gamma_n f}(x + \gamma_n a_{n-1}^*) \quad \text{and} \quad a_n^* := \frac{x + \gamma_n a_{n-1}^* - a_n}{\gamma_n}. \quad (82)$$

Then we obtain the lower bound

$$G_f(x, x^*) = f(x) + f^*(x^*) - \langle x, x^* \rangle \geq \sum_{k=1}^{\infty} \frac{\|x - \text{Prox}_{\gamma_k f}(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k}. \quad (83)$$

Proof. Because ∂f is maximally cyclically monotone, the result thus follows by combining (73), (74), and (80). $\square$

Remark 7.5. We consider again Corollary 7.4 and the notation used there. Suppose that $(x, x^*) \in \text{dom } f \times \text{dom } f^*$. Then (83) yields

$$\sum_{k=1}^{\infty} \frac{\|x - \text{Prox}_{\gamma_k f}(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k} < +\infty; \quad \text{hence,} \quad \frac{x - \text{Prox}_{\gamma_k f}(x + \gamma_k a_{k-1}^*)}{\sqrt{\gamma_k}} \rightarrow 0. \quad (84)$$

If we truncate the infinite series in (83) after the first term and set $\gamma = \gamma_1$, then we obtain Carlier's bound (see Theorem 4.3)

$$f(x) + f^*(x^*) - \langle x, x^* \rangle \geq \frac{\|x - \text{Prox}_{\gamma f}(x + \gamma x^*)\|^2}{\gamma}. \quad (85)$$

Example 7.6. Suppose that $f = \iota_U$, where U is a closed linear subspace of X . Then $\text{Prox}_{\gamma f} = P_U$ and for every $n \geq 1$, we have

$$a_n = P_U(x + \gamma_n a_{n-1}^*) = P_U x + \gamma_n P_U a_{n-1}^* \quad (86)$$

$$\text{and } a_n^* = \frac{x + \gamma_n a_{n-1}^* - a_n}{\gamma_n} = \frac{x + \gamma_n a_{n-1}^* - P_U x - \gamma_n P_U a_{n-1}^*}{\gamma_n} \quad (87a)$$

$$= \frac{P_{U^\perp} x}{\gamma_n} + P_{U^\perp} a_{n-1}^* \in U^\perp \quad (87b)$$

$$\vdots \quad (87c)$$

$$= \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2} + \cdots + \frac{1}{\gamma_n} \right) P_{U^\perp} x + P_{U^\perp} x^* \in U^\perp; \quad (87d)$$

thus, $a_1 = P_U x + \gamma_1 P_U x^*$ and $a_2 = \cdots = a_n = P_U x$. It follows that

$$\frac{\|x - \text{Prox}_{\gamma_1 f}(x + \gamma_1 a_0^*)\|^2}{\gamma_1} = \frac{\|x - P_U(x + \gamma_1 x^*)\|^2}{\gamma_1} = \frac{1}{\gamma_1} \|P_{U^\perp} x\|^2 + \gamma_1 \|P_U x^*\|^2 \quad (88)$$

and that for every $k \geq 2$

$$\begin{aligned} \frac{\|x - \text{Prox}_{\gamma_k f}(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k} &= \frac{\|x - P_U(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k} \\ &= \frac{1}{\gamma_k} \|P_{U^\perp} x\|^2 + \gamma_k \|P_U a_{k-1}^*\|^2 = \frac{1}{\gamma_k} \|P_{U^\perp} x\|^2. \end{aligned}$$

Therefore, the lower bound in (83) turns into

$$\gamma_1 \|P_U x^*\|^2 + \left(\sum_{k=1}^{\infty} \frac{1}{\gamma_k} \right) \|P_{U^\perp} x\|^2 \quad (89)$$

which is strictly larger than Carlier's bound whenever $x \notin U$.

Example 7.7. Suppose that $f = \frac{1}{2} \|\cdot\|^2 = f^*$, let $\gamma > 0$ and set $\gamma_n = \gamma$ for all $n \geq 1$. Then $\text{Prox}_{\gamma f} = (\text{Id} + \gamma \text{Id})^{-1} = (1 + \gamma)^{-1} \text{Id}$. Then for every $n \geq 1$, we have

$$a_n = \frac{x + \gamma a_{n-1}^*}{1 + \gamma} \quad (90)$$

$$\text{and } a_n^* = \frac{x + \gamma a_{n-1}^* - a_n}{\gamma} = \frac{x + \gamma a_{n-1}^* - (1 + \gamma)^{-1}(x + \gamma a_{n-1}^*)}{\gamma} \quad (91a)$$

$$= \frac{1}{1 + \gamma} x + \frac{\gamma}{1 + \gamma} a_{n-1}^* \quad (91b)$$

$$\vdots \quad (91c)$$

$$= \left(1 - \frac{\gamma^n}{(1 + \gamma)^n} \right) x + \frac{\gamma^n}{(1 + \gamma)^n} x^*. \quad (91d)$$

It follows that

$$\frac{\|x - \text{Prox}_{\gamma_1 f}(x + \gamma_1 a_0^*)\|^2}{\gamma_1} = \frac{\|x - (1 + \gamma)^{-1}(x + \gamma x^*)\|^2}{\gamma} = \frac{\gamma}{(1 + \gamma)^2} \|x - x^*\|^2 \quad (92)$$

and that for every $k \geq 2$

$$\begin{aligned} \frac{\|x - \text{Prox}_{\gamma_k f}(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k} &= \frac{\gamma}{(1 + \gamma)^2} \|x - a_{k-1}^*\|^2 \\ &= \frac{\gamma}{(1 + \gamma)^2} \frac{\gamma^{2(k-1)}}{(1 + \gamma)^{2(k-1)}} \|x - x^*\|^2 = \frac{\gamma^{2k-1}}{(1 + \gamma)^{2k}} \|x - x^*\|^2. \end{aligned}$$

That is,

$$(\forall k \geq 1) \quad \frac{\|x - \text{Prox}_{\gamma_k f}(x + \gamma_k a_{k-1}^*)\|^2}{\gamma_k} = \frac{\gamma^{2k-1}}{(1 + \gamma)^{2k}} \|x - x^*\|^2. \quad (93)$$

Therefore, the lower bound in (83) turns into

$$\left(\sum_{k=1}^{\infty} \frac{\gamma^{2k-1}}{(1 + \gamma)^{2k}} \right) \|x - x^*\|^2 = \frac{\gamma}{1 + 2\gamma} \|x - x^*\|^2 \quad (94)$$

which is strictly greater than Carlier's bound $\gamma(1 + \gamma)^{-2} \|x - x^*\|^2$ whenever $x \neq x^*$.

Acknowledgments. We thank an anonymous referee for constructive and helpful comments, and Guillaume Carlier for sending us his beautiful preprint [8]. We also would like to thank Patrick L. Combettes for suggesting [9] and [10] as references. HHB and XW were supported by NSERC Discovery Grants.

References

- [1] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, X. Wang: *Fitzpatrick functions, cyclic monotonicity and Rockafellar's antiderivative*, Nonlinear Analysis 66 (2007) 1198–1223.
- [2] H. H. Bauschke, J. M. Borwein, X. Wang: *Fitzpatrick functions and continuous linear monotone operators*, SIAM J. Optimization 18/3 (2007) 789–809.
- [3] H. H. Bauschke, P. L. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, Berlin (2017).
- [4] H. H. Bauschke, D. A. McLaren, H. S. Sendov: *Fitzpatrick functions: inequalities, examples, and remarks on a problem by S. Fitzpatrick*, J. Convex Analysis 13/3+4 (2006) 499–523.
- [5] A. Beck: *First-Order Methods in Optimization*, SIAM, Philadelphia (2017).
- [6] R. S. Burachik, M. N. Dao, S. B. Lindstrom: *Generalized Bregman envelopes and proximity operators*, J. Optimization Theory Appl. 190/3 (2021) 744–778.
- [7] R. S. Burachik, M. N. Dao, S. B. Lindstrom: *The generalized Bregman distance*, SIAM J. Optimization 31/1 (2021) 404–424.
- [8] G. Carlier: *Fenchel-Young inequality with a remainder and applications to convex duality and optimal transport*, SIAM J. Optimization (2022).

- [9] P. L. Combettes, J.-C. Pesquet: *Proximal splitting methods in signal processing*, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, et al. (eds.), Springer, New York (2011) 185–212.
- [10] P. L. Combettes, V. R. Wajs: *Signal recovery by proximal forward-backward splitting*, Multiscale Mod. Simul. 4/4 (2005) 1168–1200.
- [11] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, D. E. Knuth: *On the Lambert W function*, Adv. Comp. Math. 5/1 (1996) 329–359.
- [12] S. Fitzpatrick: *Representing monotone operators by convex functions*, in: Workshop/Miniconference on Functional Analysis and Optimization, Canberra 1988, Proc. of the Centre for Mathematical Analysis, Australian National University Vol. 20, Canberra (1988) 59–65.
- [13] B. S. Mordukhovich, N. M. Nam: *Convex Analysis and Beyond. I: Basic Theory*, Springer, Berlin (2022).
- [14] R. T. Rockafellar: *Convex Analysis*, Princeton University Press, Princeton (1970).
- [15] R. T. Rockafellar, R. J-B Wets: *Variational Analysis*, Springer, Berlin (2004).
- [16] S. Simons: *From Hahn-Banach to Monotonicity*, Springer, Berlin (2008).