

# The NC-proximal average for multiple functions

W. Hare · C. Planiden

Received: 14 January 2013 / Accepted: 5 April 2013 / Published online: 17 April 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** The NC-proximal average is a parametrized function used to continuously transform one proper, lsc, prox-bounded function into another. Until now it has been defined for two functions. The purpose of this article is to redefine it so that any finite number of functions may be used. The layout generally follows that of Hare (SIAM J Optim 20(2):650–666, 2009), extending those results to the more general case and in some instances giving alternate proofs by using techniques developed after the publication of that paper. We conclude with an example examining the discontinuity of the minimizers of the NC-proximal average.

**Keywords** Proximal average · Prox-regularity · Moreau envelope · Minima

## 1 Introduction

In 2008, Bauschke et al. [2], first addressed the question of how to transform one convex function into another in a continuous manner. Given proper convex functions  $f_0$  and  $f_1$ , their proposed solution, the *proximal average*, used Fenchel conjugates to define a parameterized function  $PA(x, \lambda)$  such that  $PA$  is epi-continuous with respect to  $\lambda$ , and  $PA(x, 0) = f_0(x)$ ,  $PA(x, 1) = f_1(x)$  for all  $x$ . The proximal average has been studied extensively since its original conception, and many favourable properties and applications of this approach have arisen [1, 3, 4, 6–9, 11, 13–15, 17, 18]. For example, the minimizers of the proximal average function change continuously with respect to  $\lambda$  [8].

---

Research by these authors was supported by UBC UGF and by NSERC of Canada.

---

W. Hare (✉) · C. Planiden  
Mathematics, University of British Columbia, 3333 University Way, Kelowna, Canada  
e-mail: warren.hare@ubc.ca

C. Planiden  
e-mail: chayneplaniden@hotmail.com

The proximal average has also been generalized and reformulated in a number of useful manners. For example, in [1], the proximal average is generalized to a finite number of convex functions. In [5], the proximal average is generalized to allow for alternate kernels, which further allowed for applications with monotone operators. In [9], the proximal average is reformulated to apply to saddle functions. And, in [11], the proximal average was reformulated to work with two (nonconvex), proper, lsc, prox-bounded functions. This document generalizes the work done in [11] to allow for a finite number of such functions.

Given two proper, lsc, prox-bounded functions,  $f_0$  and  $f_1$ , the NC-proximal average was originally defined as

$$PA_r(x, \lambda) := -e_{r+\lambda(1-\lambda)}(- (1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

where  $\lambda \in [0, 1]$  and  $e_r f$  is the Moreau envelope of  $f$  using the prox-parameter  $r$ , defined as

$$e_r f(x) := \inf_y \left\{ f(y) + \frac{r}{2}|y - x|^2 \right\}.$$

Associated with the Moreau envelope, and closely related to the NC-proximal average, is the proximal point mapping  $P_r f$  defined as

$$P_r f(x) := \operatorname{argmin}_y \left\{ f(y) + \frac{r}{2}|y - x|^2 \right\}.$$

In [11] the function  $PA_r$  is analyzed and a number of propositions and theorems are developed in order to describe its properties. Here, we extend those results for a finite number of proper, lsc, prox-bounded functions  $f_i, i \in \{1, 2, \dots, m\}$ . We begin by defining the NC-proximal average as

$$PA_{r,\delta}(x, \lambda) := -e_{r+\delta(\lambda)}\left(-\sum_{i=1}^m \lambda_i e_r f_i\right)(x), \tag{1.1}$$

$$\lambda \in \Lambda := \left\{ (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^m : \lambda_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^m \lambda_i = 1 \right\},$$

and  $\delta$  is any continuous function such that  $\delta(\lambda) = 0$  if  $\lambda = e_i$  (the canonical unit vector whose  $i^{\text{th}}$  component is 1) for some  $i$ , and  $\delta(\lambda) > 0$  otherwise. This definition generalizes that of [11] in two respects. First, the original definition is restricted to outer prox-parameter  $r + \lambda(1 - \lambda)$ , when in fact the  $\lambda(1 - \lambda)$  term can be replaced by any function  $\delta$  as described above. Second, the results found in [11] are reworked in order to accommodate any finite number of functions.

*Remark 1* It should be clear that the choice of the function  $\delta$  used in defining the NC-proximal average will have a great impact on the parameterized function  $PA_{r,\delta}$ . However, it will become clear in this paper that the underlying properties of  $PA_{r,\delta}$

are in fact not effected by  $\delta$ . As such, for ease of notation, except when necessary we shall simplify  $PA_{r,\delta}$  to  $PA_r$ .

The remainder of this article is organized as follows. Section 2 provides definitions and shows that  $PA_r$  is well-defined. Section 3 explores the prox-regularity and para-prox-regularity aspects of the function, and Sect. 4 considers its stability. We conclude, in Sect. 5, with some discussion on the minimizers of the NC-proximal average, including an example that demonstrates that the minimizers of the NC-proximal average may be multi-valued and discontinuous.

### 2 Preliminaries

Throughout this paper, we use  $q$  to represent the norm-squared function,  $q(x) = |x|^2$ . This section restates some definitions we need, and shows that under basic assumptions,  $PA_r$  is a well-defined function.

**Definition 1** A proper function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  is said to be *prox-bounded* if there exist  $r > 0$  and a point  $\bar{x}$  such that  $e_r f(\bar{x}) > -\infty$ . The infimum of the set of all such  $r$  is called the *threshold of prox-boundedness*.

**Definition 2** A function is *lower- $\mathcal{C}^2$*  on an open set  $V$  if it is finite-valued on  $V$  and at any point  $x \in V$  the function appended with a quadratic term is convex on some open convex neighborhood  $V'$  of  $x$ . The function is said to be lower- $\mathcal{C}^2$  (with no mention of  $V$ ) if  $V = \mathbb{R}^n$ .

Our first task is to confirm that  $PA_r$  is a well-defined and well-behaved function. The following proposition generalizes [11, Prop 2.5].

**Proposition 1** For  $i \in \{1, 2, \dots, m\}$  let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, prox-bounded functions with respective thresholds  $\bar{r}_i$ . Let  $r > \max_i \{\bar{r}_i\}$ . Then for all  $\lambda \in \Lambda$ ,  $PA_r$  is a proper function in  $x$ . Furthermore, if  $\lambda_i \neq 1$  for all  $i$ , then  $PA_r$  defines a lower- $\mathcal{C}^2$  function in  $x$ . Finally, if for some  $i$  one has that  $f_i + \frac{r}{2}q$  is convex, then  $PA_r(\cdot, e_i) = f_i$ .

*Proof* We know that  $-e_r f_i$  is well-defined for all  $i$ , since  $r > \bar{r}_i$  for all  $i$ . By [11, Lem 2.4], which is extendible to the case of  $m$  functions, we know that  $-\sum_{i=1}^m \lambda_i e_r f_i$  is a proper, lower- $\mathcal{C}^2$ , prox-bounded function, with threshold  $\bar{r} \leq \sum_{i=1}^m \lambda_i r = r$ . Thus the Moreau envelope of  $-\sum_{i=1}^m \lambda_i e_r f_i$  is well-defined and proper whenever the prox-parameter is greater than or equal to  $r$  (as is the case when  $\lambda \in \Lambda$ ), and it is lower- $\mathcal{C}^2$  whenever the prox-parameter is strictly greater than  $r$  (as is the case when  $\lambda \in \Lambda$  and  $\lambda_i \neq 1$  for all  $i$ ). The last statement is proved by applying [16, Ex 11.26 (d)] to  $PA_r(x, e_i) = -e_r(-e_r f_i)(x)$ . □

### 3 Prox-regularity

In this section, we wish to establish the conditions under which the function  $\sum_{i=1}^m \lambda_i e_r f_i$  is para-prox-regular, so that in Sect. 4 we may explore the stability

of  $PA_r$ . Let us recall what we mean by prox-regularity and para-prox-regularity of a function.

**Definition 3** A proper function  $f$  is *prox-regular* at a point  $\bar{x}$  for  $\bar{v} \in \partial f(\bar{x})$  if  $f$  is locally lsc at  $\bar{x}$  and there exist  $\epsilon > 0$  and  $r > 0$  such that

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{3.1}$$

whenever  $x' \neq x$ ,  $|x' - \bar{x}| < \epsilon$ ,  $|x - \bar{x}| < \epsilon$ ,  $|f(x) - f(\bar{x})| < \epsilon$ ,  $v \in \partial f(x)$ , and  $|v - \bar{v}| < \epsilon$ . We say the function is *continuously prox-regular* at  $\bar{x}$  for  $\bar{v}$  if, in addition,  $f$  is continuous as a function of  $(x, v) \in \text{gph } \partial f$  at  $(\bar{x}, \bar{v})$ . The function is said to be prox-regular at  $\bar{x}$  (with no mention of  $\bar{v}$ ) if it is prox-regular at  $\bar{x}$  for every  $\bar{v} \in \partial f(\bar{x})$ , and simply prox-regular (with no mention of  $\bar{x}$ ) if it is prox-regular at  $\bar{x}$  for every  $\bar{x} \in \text{dom } f$ .

From a graphical point of view, a prox-regular function is one that is locally bounded below by quadratics of equal curvature. Para-prox-regularity is an extension of this idea that includes an extra parameter  $\lambda$ .

**Definition 4** A proper, lsc function  $f : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R} \cup \{\infty\}$  is *parametrically prox-regular* in  $x$  at  $\bar{x}$  for  $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$  with compatible parametrization by  $\lambda$  at  $\bar{\lambda}$  (also referred to as *para-prox-regular* in  $x$  at  $(\bar{x}, \bar{\lambda})$  for  $\bar{v}$ ), with parameters  $\epsilon > 0$  and  $r > 0$ , if

$$f(x', \lambda) \geq f(x, \lambda) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{3.2}$$

whenever  $x' \neq x$ ,  $|x' - \bar{x}| < \epsilon$ ,  $|x - \bar{x}| < \epsilon$ ,  $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$ ,  $|\lambda - \bar{\lambda}| < \epsilon$ ,  $v \in \partial_x f(x, \lambda)$ , and  $|v - \bar{v}| < \epsilon$ . It is *continuously para-prox-regular* in  $x$  at  $(\bar{x}, \bar{\lambda})$  for  $\bar{v}$  if, in addition,  $f$  is continuous as a function of  $(x, \lambda, v) \in \text{gph } \partial_x f$  at  $(\bar{x}, \bar{\lambda}, \bar{v})$ . If the parameter  $\bar{\lambda}$ , the subgradient  $\bar{v}$ , or the point  $\bar{x}$  is omitted, then the para-prox-regularity of  $f$  is understood to mean for all  $\bar{\lambda} \in \text{dom } f(\bar{x}, \cdot)$ , for all  $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$ , or for all  $\bar{x} \in \text{dom } f(\cdot, \bar{\lambda})$ , respectively.

**Proposition 2** For  $i \in \{1, 2, \dots, m\}$  let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, and prox-bounded with threshold  $r_i$ . Let  $r > r_i$  for all  $i$ . Define

$$F(x, \lambda) = \begin{cases} -\sum_{i=1}^m \lambda_i e_r f_i(x), & \lambda \in \Lambda \\ \infty, & \lambda \notin \Lambda. \end{cases}$$

Then  $F$  is continuously para-prox-regular at any  $\bar{x}$ , with compatible parametrization by  $\lambda$  at any  $\bar{\lambda} \in \Lambda$ . Moreover,  $F$  is lower- $\mathcal{C}^2$  and strictly continuous, and if  $(0, y) \in \partial^\infty F(\bar{x}, \bar{\lambda})$  then  $y = 0$ .

*Proof* Since  $f_i$  is proper, lsc and prox-bounded for all  $i$ , [16, Ex 10.32] gives us that  $-e_r f_i$  is lower- $\mathcal{C}^2$  for all  $i$ . The sum of lower- $\mathcal{C}^2$  functions is lower- $\mathcal{C}^2$ , and any lower- $\mathcal{C}^2$  function is strictly continuous [16, Thm 10.31], so  $F$  is lower- $\mathcal{C}^2$  and

strictly continuous. Finally, [16, Thm 9.31] states that strict continuity of  $F$  at  $(\bar{x}, \bar{\lambda})$  is equivalent to  $\partial^\infty F(\bar{x}, \bar{\lambda}) = \{0\}$ , which gives us that  $(0, y) \in \partial^\infty F(\bar{x}, \bar{\lambda}) \Rightarrow y = 0$ . This gives us all the conditions of [10, Thm 5.7], and its conclusion is the result we seek.  $\square$

*Remark 2* The proof of [11, Lemma 3.3] can also be adapted for a longer, but more direct proof of Proposition 2.

### 4 Stability

We are now ready to explore the stability of the NC-proximal average. By Proposition 2, we can see that  $PA_r$  is the Moreau envelope of a para-prox-regular function. This allows us to take advantage of the work done in [12], where the tilt stability and full stability of Moreau envelopes and proximal mappings of para-prox-regular functions was studied.

**Theorem 1** [12, Thm 4.6] *Let  $F : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, and continuously para-prox-regular at  $(\bar{x}, \bar{\lambda})$  for  $\bar{v} \in \partial_x F(\bar{x}, \bar{\lambda})$ , with parameters  $\epsilon$  and  $r$ . Assume further that  $F$  is prox-bounded with threshold  $\rho$ , and that  $F$  satisfies the following:*

1.  $(0, y) \in \partial^\infty F(\bar{x}, \bar{\lambda}) \Rightarrow y = 0$ ,
2.  $(0, \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(0) \Rightarrow \lambda' = 0$ ,
3.  $(x', \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(v')$ ,  $v' \neq 0 \Rightarrow \langle x', v' \rangle > -\rho'|v'|^2$  for some  $\rho' > 0$ ,
4.  $\partial_x F(\bar{x}, \cdot)$  has a continuous selection  $g$  near  $\bar{\lambda}$ , with  $g(\bar{\lambda}) = \bar{v}$ .

*If  $\bar{r} > \max\{\rho, \rho', r\}$ , then there exist  $K > 0$  and a neighborhood  $\mathcal{B} = B_\delta(\bar{x} + \frac{\bar{v}}{\bar{r}}, \bar{\lambda}, \bar{r})$  such that for all  $(x, \lambda, r), (x', \lambda', r') \in \mathcal{B}$  we have that  $P_r F_\lambda(x)$  and  $P_{r'} F_{\lambda'}(x')$  are single-valued, with*

$$|P_r F_\lambda(x) - P_{r'} F_{\lambda'}(x')| \leq K|r(x - \bar{x}) - r'(x' - \bar{x}), \lambda - \lambda', r - r'|,$$

where  $F_\lambda(x) = F(x, \lambda)$ .

**Lemma 1** [11, Lem 4.4] *Suppose the function  $H : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R} \cup \{\infty\}$  is finite, single-valued, and Lipschitz continuous in  $(x, \lambda)$  near  $(\bar{x}, \bar{\lambda})$  with local Lipschitz constant  $\text{Lip } H$ . Then*

$$(0, \lambda') \in D^*H(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(0) \Rightarrow \lambda' = 0,$$

and for  $\rho > \text{Lip } H$  one has

$$(x', \lambda') \in D^*H(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(v'), \quad v' \neq 0 \Rightarrow \langle x', v' \rangle > -\rho|v'|^2.$$

The next proposition is an analog of [11, Prop 4.5], rewritten to work with a finite number of functions. The proof of [11, Prop 4.5] is easily adaptable to this setting, so we present only the key details.

**Proposition 3** For  $i \in \{1, 2, \dots, m\}$ , let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, and prox-bounded with threshold  $r_i$ . Let  $r > \max_i \{r_i\}$ , and define

$$F(x, \lambda) := - \sum_{i=1}^m \lambda_i e_r f_i(x).$$

If  $P_r f_i$  is single-valued and Lipschitz continuous for all  $i$ , then the following three properties hold:

1.  $(0, \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(0) \Rightarrow \lambda' = 0$ ,
2. for some  $\rho > 0$  we have  $(x', \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(v')$ ,  $v' \neq 0 \Rightarrow \langle x', v' \rangle > -\rho|v'|^2$ , and
3. the set-valued mapping  $\partial_x F(\bar{x}, \cdot)$  has a continuous selection  $g$  near  $\bar{\lambda}$ .

*Proof* Since  $P_r f_i$  is Lipschitz continuous, we have that  $e_r f_i \in C^{1+}$  with  $\nabla e_r f_i = r(I - P_r f_i)$  [12, Thm 2.4]. Hence,

$$\begin{aligned} \partial_x F(\bar{x}, \lambda) &= \nabla_x \left( - \sum_{i=1}^m \lambda_i e_r f_i \right) (\bar{x}, \lambda) \\ &= r \left[ \left( \sum_{i=1}^m \lambda_i P_r f_i(\bar{x}) \right) - \bar{x} \right] \end{aligned}$$

which is linear in  $\lambda$ , showing Property 3. Since  $P_r f_i$  is single-valued and Lipschitz continuous, we have  $\partial_x F(x, \lambda)$  single-valued and Lipschitz continuous. Properties 1 and 2 follow by applying Lemma 1. □

**Proposition 4** For  $i \in \{1, 2, \dots, m\}$ , let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, and prox-bounded with threshold  $r_i$ . Let  $r > \max_i \{r_i\}$ . Then  $PA_r(\cdot, \lambda) + \frac{r+\delta(\lambda)}{2}q(\cdot - \bar{x})$  is convex for any  $\bar{x}$ . Hence,  $PA_r(\cdot, \lambda)$  is lower- $C^2$ .

*Proof* Define  $F_\lambda := - \sum_{i=1}^m \lambda_i e_r f_i$ . Then

$$PA_r + \frac{r + \delta(\lambda)}{2}q = -e_{r+\delta(\lambda)}(F_\lambda) + \frac{r + \delta(\lambda)}{2}q.$$

By [16, Ex 11.26], we have

$$-e_{r+\delta(\lambda)}(F_\lambda) + \frac{r + \delta(\lambda)}{2}q = \left( F_\lambda + \frac{r + \delta(\lambda)}{2}q \right)^* ((r + \delta(\lambda))\cdot),$$

where  $f^*(x) := \sup_y \{\langle x, y \rangle - f(y)\}$  is the Fenchel conjugate as defined in [2]. This is an affine function composed with a convex function (as conjugate functions are convex), and as such it is convex. Notice that shifting the argument of  $q$  by  $\bar{x}$  only results in the addition of a linear term, as

$$q(x - \bar{x}) = q(x) + 2\langle x, \bar{x} \rangle + q(\bar{x})$$

where  $q(\bar{x})$  is constant and  $2\langle x, \bar{x} \rangle$  is linear. Hence,  $PA_r + q(\cdot - \bar{x})$  is convex. □

**Theorem 2** (Stability of  $PA_{\bar{r}}$ ) For  $i \in \{1, 2, \dots, m\}$ , let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, and prox-bounded with threshold  $r_i$ . Let  $\bar{r} > \max_i \{r_i\}$  and  $\bar{r} > \rho'$  from Theorem 1 Condition 3. Suppose that for all  $i$ ,  $P_{\bar{r}} f_i$  is single-valued and Lipschitz continuous (as is the case when  $f_i$  is prox-regular). Then  $PA_{\bar{r}}$  is well-defined and lower- $\mathcal{C}^2$ . If in addition

$$\text{Lip} \left( \sum_{i=1}^m \lambda_i P_{\bar{r}} f_i - I \right) \leq 1, \tag{4.1}$$

then for any  $\bar{\lambda}$  such that  $\delta(\bar{\lambda}) > 0$  we have

1.  $PA_{\bar{r}}(\cdot, \bar{\lambda}) \in \mathcal{C}^{1+}$  as a function of  $x$
2.  $PA_{\bar{r}}$  is locally Lipschitz continuous in  $\lambda$  near  $\bar{\lambda}$
3.  $\nabla_x PA_{\bar{r}}$  is locally Lipschitz continuous in  $\lambda$  near  $\bar{\lambda}$ .

Finally, if  $f_i + \frac{\bar{r}}{2}q$  is convex then  $PA_{\bar{r}}(\cdot, e_i) = f_i(\cdot)$ .

*Proof* Let  $F(x, \lambda) = -\sum_{i=1}^m \lambda_i e_{\bar{r}} f_i(x)$ . By Proposition 1,  $PA_{\bar{r}}$  is well-defined and finite-valued. Since  $P_{\bar{r}} f_i$  is single-valued for all  $i$ ,  $P_{\bar{r}} F$  is single-valued as well. Since  $f_i$  is proper, lsc, and prox-bounded for all  $i$ , and  $\bar{r}$  is greater than each threshold  $r_i$ , Proposition 2 gives us that  $F$  is continuously para-prox-regular at  $(\bar{x}, \bar{\lambda})$  for  $\bar{v} \in \partial_x F(\bar{x}, \bar{\lambda})$ , and that  $(0, y) \in \partial^\infty F(\bar{x}, \bar{\lambda}) \Rightarrow y = 0$ . Since  $P_{\bar{r}} f_i$  is single-valued and Lipschitz continuous for all  $i$ , we have all the conditions of [11, Prop 4.5], and therefore

1.  $(0, \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(0) \Rightarrow \lambda' = 0$
2.  $(x', \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(v'), v' \neq 0 \Rightarrow \langle x', v' \rangle > -\rho|v'|^2$  for some  $\rho > 0$
3. The mapping  $\partial_x F(\bar{x}, \cdot)$  has a continuous selection  $g$  near  $\bar{\lambda}$ .

Hence the condition  $\bar{r} > \max\{\rho, \rho', r\}$  of Theorem 1 is satisfied (recall  $r = \max_i \{r_i\}$ ). Therefore, all conditions of Theorem 1 hold, and we may assume its result. Since  $\delta \in \mathcal{C}^2$ , there exists  $\bar{K} > 0$  such that

$$|\delta(\lambda') - \delta(\lambda)| \leq \bar{K}|\lambda' - \lambda|$$

for all  $\lambda', \lambda$  near  $\bar{\lambda}$ . The rest of the proof is the same as that of [11, Thm 4.6]. □

**Corollary 1** For  $i \in \{1, 2, \dots, m\}$ , let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper and lsc such that for some  $r > 0$ ,  $f_i + \frac{r}{2}q$  is convex for all  $i$ . Then  $f_i$  is prox-regular and prox-bounded, and inequality (4.1) holds. In particular, all the conditions of Theorem 2 hold.

*Proof* Since  $f_i + \frac{r}{2}q$  is convex for all  $i$ , we have that  $f_i$  is prox-bounded and lower- $\mathcal{C}^2$ , and therefore prox-regular, for all  $i$ . Since

$$P_1 \left( f_i + \frac{r}{2}q \right) = P_{r+1} f_i,$$

by [16, Prop 12.19] we have that  $I - P_1(f_i + \frac{r}{2}q)$  is Lipschitz continuous with constant at most 1. Thus

$$\text{Lip} \left\{ \sum_{i=1}^m \lambda_i P_{r+1} f_i - I \right\} = \text{Lip} \left\{ \sum_{i=1}^m \lambda_i (I - P_{r+1} f_i) \right\} \leq \sum_{i=1}^m \lambda_i = 1.$$

This provides inequality (4.1). □

### 5 Example

In 2010, Goebel et al. presented a study of the minimizers of the proximal average function for convex functions. For convex functions  $f_i$  recall that  $-e_r \left( -\sum_{i=1}^m \lambda_i e_r f_i \right) (x)$  defined the proximal average from [2]. It was shown that

$$\Phi(\lambda) := \underset{x}{\operatorname{argmin}} -e_r \left( -\sum_{i=1}^m \lambda_i e_r f_i \right) (x)$$

is single-valued and continuous, provided that all functions are bounded below and at least one function is essentially strictly convex [8, Thm 3.8]. We next show that if  $f_i$  are convex functions, then the minimizers of the NC-proximal average coincide exactly with the minimizers of the proximal average. In particular, in this case all results from [8] hold.

**Lemma 2** *For  $i \in \{1, 2, \dots, m\}$  let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  be proper, lsc, convex, and bounded below. Let  $\lambda \in \Lambda$ , then*

$$\underset{x}{\operatorname{argmin}} P A_r(x, \lambda) = \underset{x}{\operatorname{argmin}} \sum_{i=1}^m \lambda_i e_r f_i(x) = \underset{x}{\operatorname{argmin}} -e_r \left( -\sum_{i=1}^m \lambda_i e_r f_i \right) (x).$$

*Proof* The minimizers of  $P A_r(\cdot, \lambda)$  coincide with the minimizers of its Moreau envelope  $e_{r+\delta(\lambda)} P A_r(\cdot, \lambda)$ . By [16, Ex 11.26(d)], we have that  $-e_{r+\delta(\lambda)} P A_r(x, \lambda) = \left( \sum_{i=1}^m -\lambda_i e_r f_i(x) \right)$ , so the first equality holds. The second equality appears in [8, Lem 3.2]. □

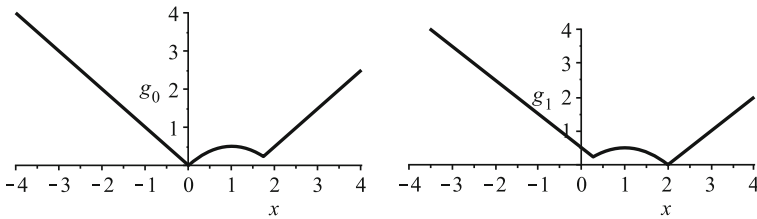
If  $f_i$  are non-convex, then the proximal average is undefined, and the results from [8] no longer apply. In this case, the results of Theorem 2 provide some small understanding of the continuity of the minimizers of the NC-proximal average, as follows.

**Corollary 2** *Let the conditions of Theorem 2 hold. Let  $x_k \in \operatorname{argmin}_x P A_r(x, \lambda_k)$ . Suppose  $\lambda_k \rightarrow \bar{\lambda}$  and  $x_k \rightarrow \bar{x}$ . Then  $\nabla P A_r(\bar{x}, \bar{\lambda}) = 0$ .*

*Proof* By Theorem 2,  $\nabla P A_r$  is Lipschitz continuous in  $\lambda$ . Therefore, there exists  $c > 0$  such that for all  $k$ ,

$$|\nabla P A_r(x_k, \lambda_k) - \nabla P A_r(x_k, \bar{\lambda})| \leq c|\lambda_k - \bar{\lambda}|.$$





**Fig. 1** Functions  $g_0$  and  $g_1$  for  $\epsilon = 0.5$

Since  $x_k \in \operatorname{argmin} P A_r(x_k, \lambda_k)$ , we know that  $\nabla P A_r(x_k, \lambda_k) = 0$ . So for all  $k$ ,

$$|\nabla P A_r(x_k, \bar{\lambda})| \leq c|\lambda_k - \bar{\lambda}|.$$

Taking the limit as  $k \rightarrow \infty$ , we find that  $\nabla P A_r(\bar{x}, \bar{\lambda}) = 0$ . □

While Corollary 2 gives us a way to identify the minimizers of  $P A_r$ , it says nothing about the single-valuedness or the continuity of said minimizers. The example that follows illustrates that, in fact, the function of minimizers of the NC-proximal average may be multi-valued and discontinuous.

Let  $\epsilon = \frac{1}{2}$ , and define the functions  $g_0$  and  $g_1$  via

$$g_0(x) := \max \left\{ -x, -\frac{1}{2}(x - 1)^2 + \frac{1}{2}, x - 2 + \epsilon \right\},$$

$$g_1(x) := \max \left\{ -x + \epsilon, -\frac{1}{2}(x - 1)^2 + \frac{1}{2}, x - 2 \right\}.$$

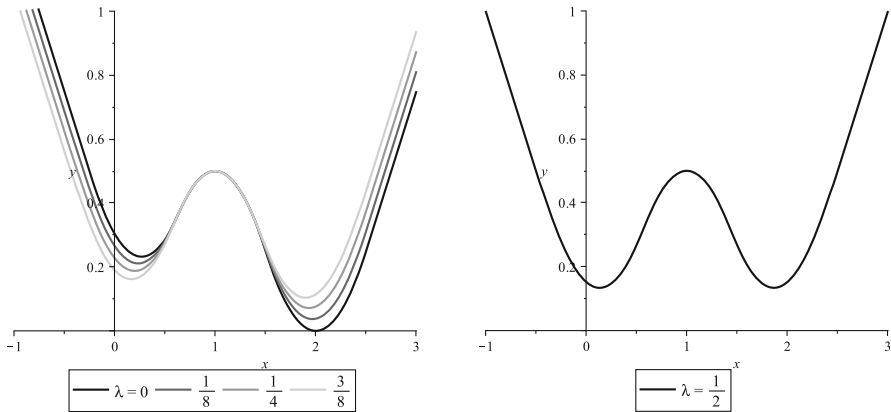
Then  $g_0$  and  $g_1$  are proper, lsc, and bounded below (see Fig. 1).

Moreover,  $g_i + \frac{1}{2}q$  is convex for  $i \in \{1, 2\}$ . Let  $k = 2 - \sqrt{4 - 2\epsilon}$ ,  $l = \sqrt{4 - 2\epsilon}$  and define

$$\begin{aligned} \delta_0 &:= 0 & \delta_1 &:= \epsilon & \epsilon_0 &:= \epsilon & \epsilon_1 &:= 0 \\ k_0 &:= 0 & k_1 &:= k & l_0 &:= l & l_1 &:= 2. \end{aligned}$$

Consider  $P_r g_i(\bar{x}) = \operatorname{argmin}_x \{g_i(x) + \frac{r}{2}|x - \bar{x}|^2\}$ . If  $r > 1$ , then we find that

$$P_r g_i(\bar{x}) = \begin{cases} \bar{x} + \frac{1}{r}, & \bar{x} < k_i - \frac{1}{r} \\ k_i, & \bar{x} \in [k_i - \frac{1}{r}, k_i - \frac{k_i}{r} + \frac{1}{r}] \\ \frac{r\bar{x}-1}{r-1}, & \bar{x} \in (k_i - \frac{k_i}{r} + \frac{1}{r}, l_i - \frac{l_i}{r} + \frac{1}{r}) \\ l_i, & \bar{x} \in [l_i - \frac{l_i}{r} + \frac{1}{r}, l_i + \frac{1}{r}] \\ \bar{x} - \frac{1}{r}, & \bar{x} > l_i + \frac{1}{r}. \end{cases}$$



**Fig. 2**  $G(\bar{x}, \lambda)$

Evaluating the Moreau envelope and simplifying, we get

$$e_r g_i(\bar{x}) = \begin{cases} -\bar{x} - \frac{1}{2r} + \delta_i, & \bar{x} < k_i - \frac{1}{r} \\ \frac{r}{2}\bar{x}^2 - rk_i\bar{x} + \frac{r-1}{2}k_i^2 + k_i, & \bar{x} \in [k_i - \frac{1}{r}, k_i - \frac{k_i}{r} + \frac{1}{r}] \\ -\frac{1}{2(r-1)}(r\bar{x}^2 - 2r\bar{x} + 1), & \bar{x} \in (k_i - \frac{k_i}{r} + \frac{1}{r}, l_i - \frac{l_i}{r} + \frac{1}{r}) \\ \frac{r}{2}\bar{x}^2 - rl_i\bar{x} + \frac{r-1}{2}l_i^2 + l_i, & \bar{x} \in [l_i - \frac{l_i}{r} + \frac{1}{r}, l_i + \frac{1}{r}] \\ \bar{x} - 2 - \frac{1}{2r} + \epsilon_i, & \bar{x} > l_i + \frac{1}{r}. \end{cases}$$

Considering the specific example  $r = 2$ , and applying  $\epsilon = \frac{1}{2}$ , we define the function  $G(\bar{x}, \lambda) := (\lambda e_2 g_0 + (1 - \lambda)e_2 g_1)(\bar{x})$ , which can be expanded to

$$G(\bar{x}, \lambda) = \begin{cases} -\bar{x} - \frac{\lambda}{2} + \frac{1}{4}, & x < -\frac{1}{2} \\ \lambda\bar{x}^2 + (\lambda - 1)\bar{x} - \frac{\lambda-1}{4}, & x \in [-\frac{1}{2}, \frac{3-2\sqrt{3}}{2}] \\ \bar{x}^2 + (\lambda - 1)(4 - 2\sqrt{3})\bar{x} - \frac{(\lambda-1)(11-6\sqrt{3})}{2}, & x \in [\frac{3-2\sqrt{3}}{2}, \frac{1}{2}] \\ (1 - 2\lambda)\bar{x}^2 + [-4 + 2\sqrt{3} + (6 - 2\sqrt{3})\lambda]\bar{x} + \frac{11-6\sqrt{3}}{2} - (6 - 3\sqrt{3})\lambda, & x \in (\frac{1}{2}, \frac{3-\sqrt{3}}{2}] \\ -\bar{x}^2 + 2\bar{x} - \frac{1}{2}, & x \in (\frac{3-\sqrt{3}}{2}, \frac{1+\sqrt{3}}{2}) \\ (2\lambda - 1)\bar{x}^2 + [2 - (2 + 2\sqrt{3})\lambda]\bar{x} - \frac{1}{2} + (2 + \sqrt{3})\lambda, & x \in [\frac{1+\sqrt{3}}{2}, \frac{3}{2}] \\ \bar{x}^2 - [4 - (4 - 2\sqrt{3})\lambda]\bar{x} + 4 - \frac{5-2\sqrt{3}}{2}\lambda, & x \in [\frac{3}{2}, \frac{1+2\sqrt{3}}{2}] \\ (1 - \lambda)\bar{x}^2 + (5\lambda - 4)\bar{x} + 4 - \frac{23}{4}\lambda, & x \in (\frac{1+2\sqrt{3}}{2}, \frac{5}{2}] \\ \bar{x} + \frac{\lambda}{2} - \frac{9}{4}, & x < \frac{5}{2}. \end{cases}$$

By Lemma 2, we know that

$$\operatorname{argmin}_{\bar{x}} PA_r(\bar{x}, \lambda) = \operatorname{argmin}_{\bar{x}} G(\bar{x}, \lambda).$$

Figure 2 displays graphs of  $G$  for various values of  $\lambda$ .

Noting that  $G \in C^1$ , we find three critical points (where  $\frac{\partial}{\partial x} G(x, \lambda) = 0$ ):

1.  $\bar{x}_1 = (1 - \lambda)(2 - \sqrt{3})$  (leftmost local minimum argument),
2.  $\bar{x}_2 = 1$  (local maximum argument),
3.  $\bar{x}_3 = 2 - (2 - \sqrt{3})\lambda$  (rightmost local minimum argument).

Observe that when  $\lambda = \frac{1}{2}$  we have that  $\bar{x}_1 = \frac{2-\sqrt{3}}{2}$ ,  $\bar{x}_3 = \frac{2+\sqrt{3}}{2}$ , and

$$G\left(\frac{2-\sqrt{3}}{2}, \frac{1}{2}\right) = \frac{2-\sqrt{3}}{2} = G\left(\frac{2+\sqrt{3}}{2}, \frac{1}{2}\right).$$

This verifies that there are two minimizers when  $\lambda = \frac{1}{2}$ . Finally, we note that

$$G(\bar{x}_1, \lambda) < G(\bar{x}_3, \lambda), \lambda \in \left[0, \frac{1}{2}\right) \quad \text{and} \quad G(\bar{x}_1, \lambda) > G(\bar{x}_3, \lambda), \lambda \in \left(\frac{1}{2}, 1\right],$$

which proves the argmin is a singleton whenever  $\lambda \neq \frac{1}{2}$ . Therefore,  $\text{argmin} PA_r$  is not a continuous function of  $\lambda$ .

## 6 Conclusion

We have seen that, using the Moreau envelope definition, the NC-proximal average can be generalized to accommodate any finite number of suitable functions. Under appropriate conditions,  $PA_r$  is well-defined, lower- $\mathcal{C}^2$ , and locally Lipschitz continuous in  $x$  and in  $\lambda$ . These properties make  $PA_r$  a useful function for researchers in the Optimization field.

## References

1. Bauschke, H., Goebel, R., Lucet, Y., Wang, X.: The proximal average: basic theory. *SIAM J. Optim.* **19**(2), 766–785 (2008). doi:[10.1137/070687542](https://doi.org/10.1137/070687542)
2. Bauschke, H., Lucet, Y., Trienis, M.: How to transform one convex function continuously into another. *SIAM Rev.* **50**(1), 115–132 (2008). doi:[10.1137/060664513](https://doi.org/10.1137/060664513)
3. Bauschke, H., Lucet, Y., Wang, X.: Primal-dual symmetric intrinsic methods for finding antiderivatives of cyclically monotone operators. *SIAM J. Control Optim.* **46**(6), 2031–2051 (2007). doi:[10.1137/060675794](https://doi.org/10.1137/060675794)
4. Bauschke, H., Moffat, S., Wang, X.: The resolvent average for positive semidefinite matrices. *Linear Algebr. Appl.* **432**(7), 1757–1771 (2010). doi:[10.1016/j.laa.2009.11.028](https://doi.org/10.1016/j.laa.2009.11.028)
5. Bauschke, H., Wang, X.: The Kernel average for two convex functions and its application to the extension and representation of monotone operators. *Trans. Am. Math. Soc.* **361**(11), 5947–5965 (2009). doi:[10.1090/S0002-9947-09-04698-4](https://doi.org/10.1090/S0002-9947-09-04698-4)
6. Bauschke, H., Wang, X., Yao, L.: Autoconjugate representers for linear monotone operators. *Math. Program.* **123**(1, Ser. B), 5–24 (2010). doi:[10.1007/s10107-009-0319-0](https://doi.org/10.1007/s10107-009-0319-0)
7. Gardiner, B., Lucet, Y.: Convex hull algorithms for piecewise linear-quadratic functions in computational convex analysis. *Set-Valued Var. Anal.* **18**(3–4), 467–482 (2010). doi:[10.1007/s11228-010-0157-5](https://doi.org/10.1007/s11228-010-0157-5)
8. Goebel, R., Hare, W., Wang, X.: The optimal value and optimal solutions of the proximal average of convex functions. *Nonlinear Anal.* **75**(3), 1290–1304 (2012). doi:[10.1016/j.na.2011.06.017](https://doi.org/10.1016/j.na.2011.06.017)
9. Goebel, R.: The proximal average for saddle functions and its symmetry properties with respect to partial and saddle conjugacy. *J. Nonlinear Convex Anal.* **11**(1), 1–11 (2010)
10. Hare, W., Planiden, C.: Parametrically prox-regular functions. Submitted to *J. Convex Anal.* (2012)

11. Hare, W.L.: A proximal average for nonconvex functions: a proximal stability perspective. *SIAM J. Optim.* **20**(2), 650–666 (2009). doi:[10.1137/07070913X](https://doi.org/10.1137/07070913X)
12. Hare, W.L., Poliquin, R.A.: Prox-regularity and stability of the proximal mapping. *J. Convex Anal.* **14**(3), 589–606 (2007)
13. Johnstone, J., Koch, V., Lucet, Y.: Convexity of the proximal average. *J. Optim. Theory Appl.* **148**(1), 107–124 (2011). doi:[10.1007/s10957-010-9747-5](https://doi.org/10.1007/s10957-010-9747-5)
14. Kum, S.: Resolvent average on second-order cone. *Taiwan. J. Math.* **15**(6), 2733–2750 (2011)
15. Lucet, Y.: What shape is your conjugate? A survey of computational convex analysis and its applications (reprint of mr2496900). *SIAM Rev.* **52**(3), 505–542 (2010). doi:[10.1137/100788458](https://doi.org/10.1137/100788458)
16. Rockafellar, R., Wets, R.B.: *Variational analysis*, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 317. Springer, Berlin (1998). doi:[10.1007/978-3-642-02431-3](https://doi.org/10.1007/978-3-642-02431-3)
17. Wang, X.: Self-dual regularization of monotone operators via the resolvent average. *SIAM J. Optim.* **21**(2), 438–462 (2011). doi:[10.1137/100795942](https://doi.org/10.1137/100795942)
18. Wang, X., Bauschke, H.: Compositions and averages of two resolvents: relative geometry of fixed points sets and a partial answer to a question by C. Byrne. *Nonlinear Anal.* **74**(13), 4550–4572 (2011). doi:[10.1016/j.na.2011.04.024](https://doi.org/10.1016/j.na.2011.04.024)