SIAM J. OPTIM.
Vol. 29, No. 2, pp. 1366–1391

POLAR CONVOLUTION*

MICHAEL P. FRIEDLANDER†, IVES MACÊDO‡, AND TING KEI PONG§

Abstract. The Moreau envelope is one of the key convexity-preserving functional operations in convex analysis, and it is central to the development and analysis of many approaches for convex optimization. This paper develops the theory for an analogous convolution operation, called the polar envelope, specialized to gauge functions. Many important properties of the Moreau envelope and the proximal map are mirrored by the polar envelope and its corresponding proximal map. These properties include smoothness of the envelope function, uniqueness, and continuity of the proximal map, which play important roles in duality and in the construction of algorithms for gauge optimization. A suite of tools with which to build algorithms for this family of optimization problems is thus established.

Key words. gauge optimization, max convolution, proximal algorithms

AMS subject classifications. 90C15, 90C25

DOI. 10.1137/18M1209088

1. Motivation.

Let f_1 and f_2 be proper convex functions that map a finite-dimensional Euclidean space $\mathcal{X}$ to the extended reals $\mathbb{R} \cup \{+\infty\}$. The *infimal sum convolution* operation

$$(f_1 \,\square\, f_2)(x) = \inf_z \{\, f_1(z) + f_2(x - z) \,\} \tag{1.1}$$

results in a convex function that is a "mixture" of f_1 and f_2. As is usually the case in convex analysis, the operation is best understood from the epigraphical viewpoint: if the infimal operation attains its optimal value at each x, then the resulting function $f_1 \,\square\, f_2$ satisfies the relationship

$$\operatorname{epi}(f_1 \,\square\, f_2) = \operatorname{epi} f_1 + \operatorname{epi} f_2.$$

Sum convolution is thus also known as *epigraphical addition* [14, Page 34].

This operation emerges in various forms in convex optimization. Take, for example, the function $f_2 = \frac{1}{2\alpha}\|\cdot\|_2^2$ for some positive scalar α. Then

$$\left(f_1 \,\square\, \tfrac{1}{2\alpha}\|\cdot\|_2^2\right)(x) = \inf_z \left\{ f_1(z) + \tfrac{1}{2\alpha}\|x - z\|_2^2 \right\} \tag{1.2}$$

is the Moreau envelope of f_1, and the minimizing set

$$\mathsf{prox}_{\alpha f_1}(x) := \arg\min_z \left\{ f_1(z) + \tfrac{1}{2\alpha}\|x - z\|_2^2 \right\}$$

*Received by the editors August 22, 2018; accepted for publication (in revised form) March 4, 2019; published electronically May 16, 2019.
http://www.siam.org/journals/siopt/29-2/M120908.html
Funding: The first author is supported by ONR award N00014-17-1-2009. The third author is supported by Hong Kong Research Grants Council award PolyU253008/15p.
†Department of Computer Science and Department of Mathematics, University of British Columbia, 2366 Main Mall, Vancouver, BC, V6T 1Z4, Canada (mpf@cs.ubc.ca).
‡Current address: Amazon, 510 West Georgia St., Vancouver, BC, V6B 0M3, Canada (research@ivesmacedo.com).
§Department of Applied Mathematics, the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (tk.pong@polyu.edu.hk).

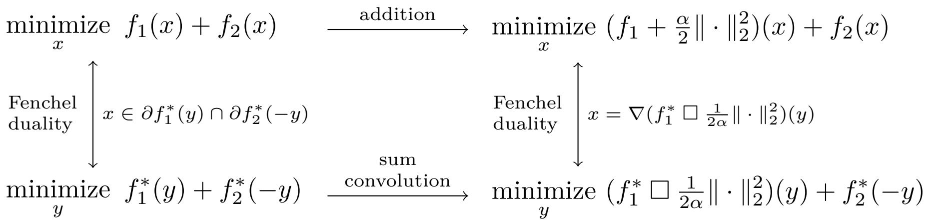

Fig. 1.1. *The Fenchel duality correspondence for regularized problems. The formulas on the vertical arrows show how a primal solution x may be recovered from a dual solution y* [15, Example 11.41]. *Strong convexity of the regularized problem ensures that x can be recovered uniquely.*

is the proximal map of αf_1. The proximal map features prominently in first-order methods for minimizing nonsmooth convex functions, where it appears as a key computational kernel. For example, the proximal map is required at each iteration of a splitting method, including the proximal-gradient and Douglas–Rachford algorithms, often used in practice. Bauschke and Combettes [4, Chapter 28] describe this class of methods in detail.

Sum convolution also appears naturally in the process of dualization through conjugacy. The close relationship between sum convolution and conjugacy is encapsulated by the formula

$$(f_1 \,\square\, f_2)^* = f_1^* + f_2^*, \tag{1.3}$$

which reveals the duality between sum convolution and addition [14, Theorem 16.4]. This fact is often leveraged as a device to solve nonsmooth problems via algorithms that are restricted to differentiable functions, as follows: the objective function is regularized, which results, via conjugacy, in a dual objective that is the sum convolution of the conjugated regularized objective. Under the appropriate conditions, Fenchel duality provides the necessary correspondence, as illustrated in Figure 1.1. The key property of the regularized dual objective $(f_1^* \,\square\, \frac{1}{2\alpha}\|\cdot\|_2^2)$, shown in the bottom right of this figure, is that its gradient is $\frac{1}{\alpha}$-Lipschitz continuous [4, Proposition 12.29]. A variety of first-order algorithms can then be applied to the regularized dual problem in order to approximate a solution of the regularized primal. This approach is often used in practice. For example, it forms the backbone for the TFOCS software package [6] and for Nesterov's smoothing algorithm and its variants [5, 13].

The approach illustrated in Figure 1.1 is fully general, in the sense that it applies to all of convex optimization. However, such generality may unravel useful structures already present in the original problem. Homogeneity, in particular, is not preserved under the Moreau envelope. This property often appears in sparse optimization problems, and more generally, in gauge optimization problems, which involve nonnegative sublinear functions [9]. The smoothing approach described in Figure 1.1 does not apply in the gauge setting: even if the problems on the left-hand side of the figure are gauge problems, the functions on the right-hand side cannot be gauges, because the regularization is not homogeneous. We are thus led to consider an alternative convexity-preserving convolution operation that can be used to construct smooth approximations and that appears naturally as part of the dualization process.

Infimal max convolution is a convexity-preserving transformation similar to infimal sum convolution. However, the sum operation in (1.1) is replaced by a "max" between the two functions:

$$(f_1 \diamond f_2)(x) = \inf_z \max \{ f_1(z),\ f_2(x-z) \}.$$

Max convolution first appeared in Rockafellar [14, Theorem 5.8] as an example of a convexity-preserving operation, and was subsequently studied in detail by Seeger and Volle [17]. As with sum convolution, this operation mixes f_1 and f_2, except that it results in a function whose level sets are the sums of the level sets of f_1 and f_2. Seeger and Volle thus refer to max convolution as *level-set addition*.

Our main purpose in this paper is to develop further the theory of the max-convolution operation, particularly when applied to gauges. As we will describe, this operation, coupled with gauge duality, leads to a correspondence that is completely analogous to the transformations shown in Figure 1.1.

1.1. Contributions. The duality between conjugacy and sum convolution, encapsulated in (1.3), is one of the main ingredients necessary for deriving the smooth dual approach described by Figure 1.1. In section 3 we derive an analogous identity for max convolution for gauges. When κ_1 and κ_2 are gauges,

$$(\kappa_1 \diamond \kappa_2)^\circ = \kappa_1^\circ + \kappa_2^\circ.$$

Here, the map $\kappa \mapsto \kappa^\circ$ is the polar transform of the gauge κ. This identity shows that, for gauge functions, polarity enjoys a relationship with max convolution that is analogous to the relationship between Fenchel conjugation and sum convolution. Because many useful properties related to polarity accrue when specializing max convolution to gauge functions, we term the operation *polar convolution*.

A case of special importance is the polar convolution of a gauge with the 2-norm, rather than the squared 2-norm used in (1.2). Under mild conditions,

$$\left(\kappa \diamond \tfrac{1}{\alpha}\|\cdot\|_2\right)(x) = \inf_z \max\left\{\kappa(z),\ \tfrac{1}{\alpha}\|x-z\|_2\right\}$$

is smooth at x, as we prove in section 4. We dub this function the *polar envelope* of κ. In that section, we describe the subdifferential properties of the polar envelope, along with properties of the corresponding polar proximal map. In section 5 we demonstrate how to compute these quantities for several important examples.

Consider again the duality correspondence shown in Figure 1.1. In section 6 we derive an analogous correspondence that replaces Fenchel duality with gauge duality, and sum convolution with max convolution. These relationships are summarized by Figure 1.2. In this figure, κ is a closed gauge, $\mathcal{C}$ is a closed convex set that does not contain the origin, and $\mathcal{C}' := \{ y \mid \langle x, y\rangle \geq 1 \}$ is the antipolar set of $\mathcal{C}$. As is the case with Fenchel duality, the regularized dual problem is now smooth (cf. section 4), and thus first-order methods may be applied to this problem. The inclusion on the left-hand side of Figure 1.2 follows from Aravkin et al. [1, Corollary 3.7].

As another illustrative example of how the polar envelope might be applied, we consider in section 7 how the polar envelope can be used to develop proximal-point-like algorithms for convex optimization problems.

1.2. Notation. Throughout this paper, the 2-norm of a vector x is denoted by $\|x\|_2 = \sqrt{\langle x, x\rangle}$, but we often simply denote it by $\|x\|$. The α-level set for a function f is denoted by $[f \leq \alpha] = \{ x \mid f(x) \leq \alpha \}$. For a convex set $\mathcal{C}$, let $\mathsf{dist}_{\mathcal{C}}(x) := \inf \{ \|x-z\| \mid z \in \mathcal{C} \}$ denote the Euclidean distance of x to the set $\mathcal{C}$, and $\mathsf{proj}_{\mathcal{C}}(x) := \arg\min \{ \|x-z\| \mid z \in \mathcal{C} \}$ denote the corresponding projection. The recession cone of

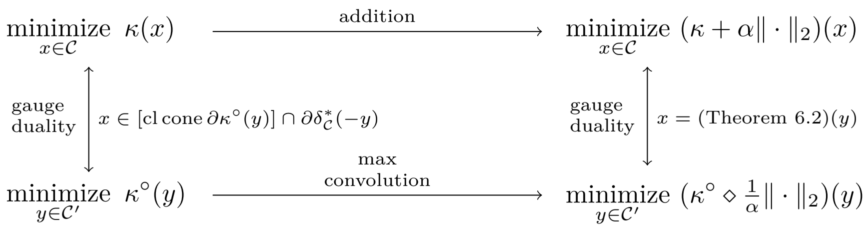

FIG. 1.2. *The gauge duality correspondence for regularized gauge problems. The formulas on the vertical arrows show how a primal solution x may be recovered from a dual solution y. Under gauge duality, the homogeneous regularizer ensures that x can be recovered uniquely.*

$\mathcal{C}$ is denoted by $\mathcal{C}_\infty$; its polar is denoted by $\mathcal{C}^\circ := \{\, u \mid \langle u, x\rangle \le 1 \ \forall x \in \mathcal{C} \,\}$. When $\mathcal{C}$ is a cone, its polar simplifies to the closed cone $\mathcal{C}^\circ = \{\, u \mid \langle u, x\rangle \le 0 \ \forall x \in \mathcal{C} \,\}$. Finally, fractions such as $(1/(2\alpha))$ are often abbreviated as $(1/2\alpha)$.

2. The max-convolution operation.

In this section we review the max-convolution operation, and establish a baseline set of results for later sections. Max convolution is also known as a *level sum* [17] because $(f_1 \diamond f_2)$ is the sum of the strict level sets of f_1 and f_2:

$$[(f_1 \diamond f_2) < \lambda] = [f_1 < \lambda] + [f_2 < \lambda] \quad \forall \lambda \in \mathbb{R}; \tag{2.1}$$

see Rockafellar [14, Page 40]. This property mirrors that for sum convolution, where instead it is the strict epigraphs that are summed:

$$\{\, (x, \alpha) \mid (f_1 \,\Box\, f_2)(x) < \alpha \,\} = \{\, (x_1, \alpha_1) \mid f_1(x_1) < \alpha_1 \,\} + \{\, (x_2, \alpha_2) \mid f_2(x_2) < \alpha_2 \,\}.$$

In both cases, if the infimal operations that define these convolutions are attained, then the convolution is *exact*, and all of the strict inequalities become weak. Seeger [16] uses the term *inverse sum* to describe the convolution $(f_1 \diamond f_2)$ when f_1 and f_2 are continuous nonnegative sublinear functions. We follow Zălinescu [19, Theorem 2.1.3(ix)] and adopt the term "max convolution" to highlight its convolutional nature.

The perspective transform for any proper convex function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ creates a convex function whose epigraph is $\operatorname{cone}(\operatorname{epi} f \times \{1\})$, which is the cone generated by the "lifted" set. The perspective transform does not necessarily preserve closure, and so it is convenient for us to redefine a closed version by

$$f^\pi(x, \lambda) := \begin{cases} \lambda f\left(\lambda^{-1}x\right) & \text{if } \lambda > 0, \\ f_\infty(x) & \text{if } \lambda = 0, \\ \infty & \text{if } \lambda < 0, \end{cases} \tag{2.2}$$

where f_∞ is the recession function of f [3, Definition 2.5.1]. The perspective f^π of a proper closed convex function f is proper closed convex [14, Page 67].

Seeger and Volle [17] derive an expression for the conjugate of the max convolution. (Seeger and Volle attribute this result to Attouch [2] and Zălinescu [18].) Here we provide a slightly rephrased version of that result expressed in terms of the perspective function. We adopt the notation

$$(\lambda^+ f)(x) := \begin{cases} \lambda f(x) & \text{if } \lambda > 0, \\ \delta_{\operatorname{dom} f}(x) & \text{if } \lambda = 0, \end{cases}$$

which emphasizes the role of the limiting behavior as $\lambda \to 0$ from the right. Thus,

$$\partial(\lambda^+ f)(x) = \begin{cases} \{\, \lambda z \mid z \in \partial f(x) \,\} & \text{if } \lambda > 0, \\ \mathcal{N}(x \mid \operatorname{dom} f) & \text{if } \lambda = 0. \end{cases}$$

Proposition 2.1 (Seeger and Volle [17, Proposition 4.1]). *Let f_1 and f_2 be proper convex functions. Then*

$$(f_1 \diamond f_2)^*(y) = \inf \{\, (f_1^*)^\pi(y, \lambda_1) + (f_2^*)^\pi(y, \lambda_2) \mid \lambda_1 + \lambda_2 = 1,\ \lambda_i \geq 0 \,\}.$$

Proof. Seeger and Volle [17, Proposition 4.1] show that

$$(f_1 \diamond f_2)^*(y) = \inf \{\, ([\lambda_1^+ f_1]^* + [\lambda_2^+ f_2]^*)(y) \mid \lambda_1 + \lambda_2 = 1,\ \lambda_i \geq 0 \,\}. \tag{2.3}$$

For a proper convex function g and any nonnegative scalar λ,

$$(\lambda^+ g)^*(y) = \begin{cases} \lambda g^*(\lambda^{-1} y) & \text{if } \lambda > 0, \\ \delta^*_{\operatorname{dom} g}(y) & \text{if } \lambda = 0. \end{cases}$$

The desired conclusion thus follows from (2.3), from the definition of perspective functions, and from Auslender and Teboulle [3, Theorem 2.5.4(a)], which asserts $\delta^*_{\operatorname{dom} g} = (g^*)_\infty$. ◻

We also recall the following formula for the subdifferential of the max convolution $f_1 \diamond f_2$, originally given by Seeger and Volle [17, Proposition 4.3]. This result will be useful in establishing the differential properties of the max convolution applied to gauge functions; cf. sections 4 and 7.

Proposition 2.2 (Seeger and Volle [17]). *Let f_1 and f_2 be proper closed convex functions. Let $\bar{x} \in \operatorname{dom}(f_1 \diamond f_2)$ and suppose $(f_1 \diamond f_2)(\bar{x}) = \max\{\, f_1(\bar{x}_1),\ f_2(\bar{x}_2) \,\}$ for some $\bar{x}_1 + \bar{x}_2 = \bar{x}$, with $\bar{x}_1 \in \operatorname{dom} f_1$ and $\bar{x}_2 \in \operatorname{dom} f_2$. Then*

$$\partial(f_1 \diamond f_2)(\bar{x}) = \bigcup \{\, \partial(\mu^+ f_1)(\bar{x}_1) \cap \partial([1-\mu]^+ f_2)(\bar{x}_2) \;\big|\; \mu \in \Gamma \,\},$$

where $\Gamma = \{\, \mu \in [0,1] \mid \mu^+ f_1(\bar{x}_1) + (1-\mu)^+ f_2(\bar{x}_2) = (f_1 \diamond f_2)(\bar{x}) \,\}$.

3. Polar convolution.

Any nonnegative convex function that is positively homogeneous and zero at the origin is called a *gauge*. This family of functions includes, for example, all norms and seminorms, and any support function on a set whose convex hull includes the origin. When max convolution is specialized to gauge functions, the analogues between the max convolution and the sum convolution deepen. Here is our formal definition of polar convolution.

Definition 3.1 (polar convolution). *Let κ_1 and κ_2 be gauges. Their* polar convolution *is defined by $\kappa_1 \diamond \kappa_2$.*

Although the max convolution of two general proper convex functions can be improper (i.e., it may attain $-\infty$), the polar convolution of two gauges is necessarily a proper convex function. Indeed, one can show that $\kappa_1 \diamond \kappa_2$ is a gauge. Moreover, we make the following immediate observation, whose simple proof is omitted.

Proposition 3.2 (lower approximation). *Let κ_1 and κ_2 be gauges. Then $\kappa_1 \diamond \kappa_2$ is a gauge and $(\kappa_1 \diamond \kappa_2)(x) \leq \min\{\kappa_1(x),\ \kappa_2(x)\}$ for all $x \in \mathcal{X}$. In particular, if either $\operatorname{dom} \kappa_1 = \mathcal{X}$ or $\operatorname{dom} \kappa_2 = \mathcal{X}$, then $\operatorname{dom}(\kappa_1 \diamond \kappa_2) = \mathcal{X}$.*

In the same way that every convex function can be paired with its Fenchel conjugate, every gauge function κ can be paired with its polar, defined by

$$\kappa^\circ(y) = \inf\{\, \mu > 0 \mid \langle x, y\rangle \le \mu\kappa(x)\ \forall x \,\}. \tag{3.1}$$

This leads to the polar-gauge inequality

$$\langle x, y\rangle \le \kappa(x)\cdot\kappa^\circ(y) \quad \forall x \in \operatorname{dom}\kappa,\ \forall y \in \operatorname{dom}\kappa^\circ, \tag{3.2}$$

and thus the polar κ° is the function that satisfies this inequality most tightly. The following lemma reveals the duality between polar convolution and addition under the polarity operation.

LEMMA 3.3 (polar convolution identity). *Let κ_1 and κ_2 be gauges. Then*

$$(\kappa_1 \diamond \kappa_2)^\circ = \kappa_1^\circ + \kappa_2^\circ. \tag{3.3}$$

If either κ_1 or κ_2 is also continuous, then

$$\kappa_1 \diamond \kappa_2 = (\kappa_1^\circ + \kappa_2^\circ)^\circ.$$

Proof. It follows directly from the definition (3.1) of a polar gauge κ that

$$\kappa^\circ(y) = \sup_x \{\, \langle x, y\rangle \mid \kappa(x) \le 1 \,\}$$

for any y. Thus, a direct computation shows that, for any y,

$$\begin{aligned}
(\kappa_1 \diamond \kappa_2)^\circ(y) &= \sup_x \{\, \langle x, y\rangle \mid (\kappa_1 \diamond \kappa_2)(x) \le 1 \,\} \overset{\text{(i)}}{=} \sup_x \{\, \langle x, y\rangle \mid (\kappa_1 \diamond \kappa_2)(x) < 1 \,\}\\
&= \sup_{x,\, x_1} \{\, \langle x, y\rangle \mid \kappa_1(x_1) < 1,\ \kappa_2(x - x_1) < 1 \,\}\\
&= \sup_{x,\, x_1} \{\, \langle x - x_1 + x_1, y\rangle \mid \kappa_1(x_1) < 1,\ \kappa_2(x - x_1) < 1 \,\}\\
&\overset{\text{(ii)}}{=} \sup_{x_1,\, x_2} \{\, \langle x_1 + x_2, y\rangle \mid \kappa_1(x_1) \le 1,\ \kappa_2(x_2) \le 1 \,\}\\
&= \kappa_1^\circ(y) + \kappa_2^\circ(y),
\end{aligned}$$

where equalities (i) and (ii) follow from Rockafellar [14, Theorem 7.6] and the continuity of the linear function $x \mapsto \langle x, y\rangle$. Now, if in addition either κ_1 or κ_2 is continuous, then we see from Proposition 3.2 that $\kappa_1 \diamond \kappa_2$ is continuous. The second conclusion now follows immediately by taking the polar of both sides of (3.3) and applying [9, Proposition 2.1(ii)] to the continuous gauge $\kappa_1 \diamond \kappa_2$. ∎

3.1. Polar convolution as a sum of sets. Polar convolution can also be viewed as a function induced by a convex set that defines its unit level set. This connection is most transparent when viewing gauges via their representation as a Minkowski functional $\gamma_{\mathcal{D}}$ for some nonempty convex set $\mathcal{D}$:

$$\kappa(x) = \gamma_{\mathcal{D}}(x) := \inf\{\, \lambda \ge 0 \mid x \in \lambda\mathcal{D} \,\}.$$

We first need the following result, which relates the sum of Minkowski functions of two sets to the Minkowski function of the sum of the sets.

LEMMA 3.4. *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be closed convex sets containing the origin. Then*

$$\gamma_{\mathcal{D}_1} + \gamma_{\mathcal{D}_2} = \gamma_{(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)^\circ}. \tag{3.4}$$

If additionally $0 \in \operatorname{ri}(\mathcal{D}_1 - \mathcal{D}_2)$, then $(\gamma_{\mathcal{D}_1} + \gamma_{\mathcal{D}_2})^\circ = \gamma_{\mathcal{D}_1^\circ + \mathcal{D}_2^\circ}$.

Proof. Theorem 14.5 of [14] contains most of the tools needed. In particular, the gauge of any closed convex set containing the origin is the support function of the polar. Thus,

$$\gamma_{\mathcal{D}_1} + \gamma_{\mathcal{D}_2} = \delta^*_{\mathcal{D}_1^\circ} + \delta^*_{\mathcal{D}_2^\circ}.$$

Next, observe that

$$\delta^*_{\mathcal{D}_1^\circ} + \delta^*_{\mathcal{D}_2^\circ} = \delta^*_{\mathcal{D}_1^\circ + \mathcal{D}_2^\circ} = \delta^*_{\operatorname{cl}(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)} = \delta^*_{(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)^{\circ\circ}},$$

where the second equality holds because the support function does not distinguish a set from its closure, and the last equality follows from the bipolar theorem [14, Theorem 14.5]. Again using the polarity correspondence between gauge and support functions, $\gamma_{\mathcal{D}_1} + \gamma_{\mathcal{D}_2} = \gamma_{(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)^\circ}$, as required.

We now prove the second part of the lemma. From (3.4) and Rockafellar [14, Theorem 15.1], we deduce that

$$(\gamma_{\mathcal{D}_1} + \gamma_{\mathcal{D}_2})^\circ = \gamma_{(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)^{\circ\circ}}. \tag{3.5}$$

Next, note that for any closed convex set $\mathcal{D}$ that contains the origin one has

$$\operatorname{dom} \delta^*_{\mathcal{D}^\circ} = \bigcup_{\lambda > 0} \{\, x \mid \delta^*_{\mathcal{D}^\circ}(x) \le \lambda \,\} = \bigcup_{\lambda > 0} \lambda \mathcal{D}^{\circ\circ} = \bigcup_{\lambda > 0} \lambda \mathcal{D}, \tag{3.6}$$

where the last equality follows the bipolar theorem [14, Theorem 14.5]. Take the relative interior on both sides of (3.6) and use [14, Page 50] to deduce that $\operatorname{ri}\mathcal{D} \subseteq \operatorname{ri}\operatorname{dom} \delta^*_{\mathcal{D}^\circ}$. Thus, from the assumption $0 \in \operatorname{ri}(\mathcal{D}_1 - \mathcal{D}_2)$, we obtain that

$$0 \in \operatorname{ri}(\mathcal{D}_1 - \mathcal{D}_2) = \operatorname{ri}\mathcal{D}_1 - \operatorname{ri}\mathcal{D}_2 \subseteq \operatorname{ri}(\operatorname{dom} \delta^*_{\mathcal{D}_1^\circ} - \operatorname{dom} \delta^*_{\mathcal{D}_2^\circ}), \tag{3.7}$$

where the equality follows from [14, Corollary 6.6.2]. Now, the relation (3.7), together with [14, Theorem 23.8], implies that

$$\mathcal{D}_1^\circ + \mathcal{D}_2^\circ = \partial \delta^*_{\mathcal{D}_1^\circ}(0) + \partial \delta^*_{\mathcal{D}_2^\circ}(0) = \partial(\delta^*_{\mathcal{D}_1^\circ} + \delta^*_{\mathcal{D}_2^\circ})(0),$$

where we used the fact that $\mathcal{D} = \partial \delta^*_{\mathcal{D}}(0)$ for any closed convex set $\mathcal{D}$ [14, Theorem 13.1]. This proves that $\mathcal{D}_1^\circ + \mathcal{D}_2^\circ$ is closed. Thus, $(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)^{\circ\circ} = \mathcal{D}_1^\circ + \mathcal{D}_2^\circ$ by the bipolar theorem [14, Theorem 14.5]. It then follows from (3.5) that $(\gamma_{\mathcal{D}_1} + \gamma_{\mathcal{D}_2})^\circ = \gamma_{(\mathcal{D}_1^\circ + \mathcal{D}_2^\circ)}$, as required. ∎

When $\mathcal{D}_1$ and $\mathcal{D}_2$ are closed convex sets containing the origin with $0 \in \operatorname{int} \mathcal{D}_1^\circ$—which implies that $\mathcal{D}_1$ is bounded—Lemma 3.4 allows us to express the polar convolution of two gauge functions $\gamma_{\mathcal{D}_1}$ and $\gamma_{\mathcal{D}_2}$ as

$$\operatorname{cl}(\gamma_{\mathcal{D}_1} \diamond \gamma_{\mathcal{D}_2}) = (\gamma_{\mathcal{D}_1^\circ} + \gamma_{\mathcal{D}_2^\circ})^\circ = \gamma_{\mathcal{D}_1 + \mathcal{D}_2},$$

where the first equality follows from Lemma 3.3 and Friedlander, Macêdo, and Pong [9, Propositions 2.1(ii) and 2.3(i)], and the second equality follows from Lemma 3.4 and the bipolar theorem [14, Theorem 14.5]. Thus, we observe that the polar convolution of the Minkowski functions of two sets results in the Minkowski function of the sum of the sets. This result confirms the level-set addition property described by the identity (2.1).

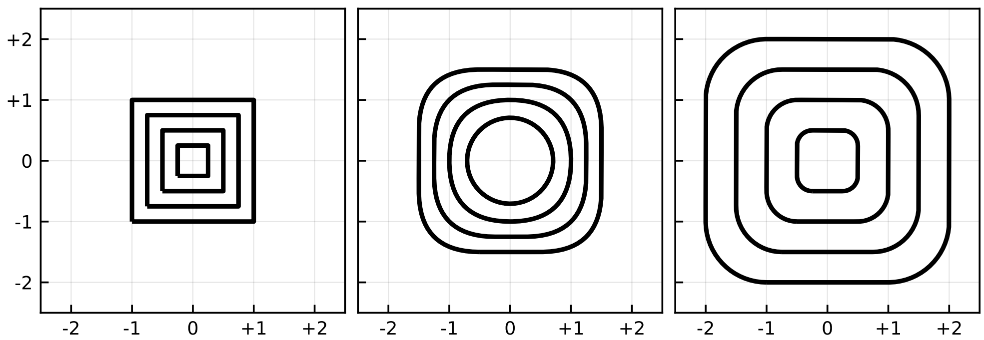

Fig. 4.1. *The* $\{.25, .50, .75, 1.0\}$ *iso-contours of the infinity norm (left), its Moreau envelope (middle), and its polar envelope (right). The iso-contours of the original norm and its polar envelope are evenly spaced because both functions are positively homogeneous, unlike the Moreau envelope.*

4. Polar envelope.

We now define a special case of polar convolution in which one of the functions involved is a multiple of the 2-norm. This operation is analogous to the *Moreau envelope* of a general convex function f; cf. (1.2). Seeger and Volle [17, Example 3.1] briefly describes the max convolution of a general convex function with the (unsquared) 2-norm. When restricted to gauge functions, the max-convolution operation has a number of useful properties that can be characterized explicitly. These include, for example, its differential properties. The formal definition of the polar envelope of a gauge function is as follows.

Definition 4.1 (polar envelope). *For a gauge* κ *and positive scalar* α*, the function*

$$\kappa_\alpha(x) := (\kappa \diamond (1/\alpha)\|\cdot\|)(x) = \inf_z \max\{\kappa(z), (1/\alpha)\|x - z\|\} \tag{4.1}$$

is the polar envelope *of* κ*. The corresponding* polar proximal map

$$\mathsf{pprox}_{\alpha\kappa}(x) := \arg\min_z \max\{\kappa(z), (1/\alpha)\|x - z\|\}$$

is the minimizing set that defines κ_α*.*

Figure 4.1 shows the Moreau and polar envelopes of the infinity norm.

Example 4.2 (indicator to a cone). Let $\mathcal{K}$ be a closed convex cone. Then $\kappa := \delta_\mathcal{K}$ is a closed gauge. Moreover, for any $\alpha > 0$, we have

$$\kappa_\alpha(x) = \inf_z \max\{\delta_\mathcal{K}(z), (1/\alpha)\|x - z\|\} = \inf_{z\in\mathcal{K}} (1/\alpha)\|x - z\| = (1/\alpha)\,\mathsf{dist}_\mathcal{K}(x).$$

For comparison, the Moreau envelope of $\delta_\mathcal{K}$ is related to the squared distance to the cone:

$$\big(\delta_\mathcal{K} \,\square\, (1/2\alpha)\|\cdot\|^2\big)(x) = (1/2\alpha)\,\mathsf{dist}^2_\mathcal{K}(x).$$

The solutions of both the max and sum convolutions are the same in this case, and correspond to the projection of x onto the cone:

$$\mathsf{pprox}_{\alpha\kappa}(x) = \mathsf{proj}_\mathcal{K}(x) = \mathsf{prox}_{\alpha\kappa}(x). \qquad \square$$

We collect some known properties of the polar envelope in the next proposition, which specializes results established by Proposition 3.2 and Corollary 4.1 of Seeger and Volle [17].

PROPOSITION 4.3 (Seeger and Volle [17]). *Let κ and ρ be gauge functions. The following properties hold:*

(i) *for any $\alpha > 0$, κ_α is Lipschitz continuous with modulus $1/\alpha$;*
(ii) $\operatorname{cl}\kappa(x) = \sup\{\,\kappa_\alpha(x) \mid \alpha > 0\,\}$ *for all x;*
(iii) *if κ and ρ are both closed, then $\kappa = \rho$ if and only if there exists $\alpha > 0$ so that $\kappa_\alpha = \rho_\alpha$.*

The following results establish the differential properties of the polar envelope and the corresponding polar proximal map.

THEOREM 4.4 (differential properties). *For any gauge κ and positive scalar α, the following properties hold.*

(i) *The subdifferential of the polar envelope κ_α is given by*

$$\partial\kappa_\alpha(x) = \arg\max_y \{\langle x, y\rangle \mid \kappa^\circ(y) + \alpha\|y\| \le 1\}.$$

Moreover, $\kappa_\alpha(x) = \langle x, y\rangle$ for any $y \in \partial\kappa_\alpha(x)$.

(ii) *If $\bar{x} \in \mathsf{pprox}_{\alpha\kappa}(x)$, then $\|x - \bar{x}\| \ge \alpha\kappa(\bar{x})$. If in addition κ is continuous, then $\|x - \bar{x}\| = \alpha\kappa(\bar{x})$.*
(iii) *If κ is closed, then $\mathsf{pprox}_{\alpha\kappa}(x)$ is a singleton for all x, and $\mathsf{pprox}_{\alpha\kappa}$ is a continuous and positively homogeneous map.*
(iv) *Suppose that κ is closed. Then κ_α is differentiable at all x such that $\kappa_\alpha(x) > 0$. Moreover, at these x, it holds that $\langle x, x - \bar{x}\rangle > 0$ and*

$$\nabla\kappa_\alpha(x) = \frac{\|x - \bar{x}\|}{\alpha\langle x, x - \bar{x}\rangle}(x - \bar{x}),$$

where $\bar{x} = \mathsf{pprox}_{\alpha\kappa}(x)$.

Proof. (i) Because κ_α is continuous, we have from Friedlander, Macêdo, and Pong [9, Proposition 2.1(ii) and (iii)] that

$$\begin{aligned}\kappa_\alpha(x) &= (\kappa_\alpha)^{\circ\circ}(x) = \sup_x \{\,\langle x, y\rangle \mid (\kappa_\alpha)^\circ(y) \le 1\,\}\\ &= \sup_x \{\,\langle x, y\rangle \mid \kappa^\circ(y) + \alpha\|y\| \le 1\,\},\end{aligned} \tag{4.2}$$

where the last equality follows from Lemma 3.3. Since the set $\{\,y \mid \kappa^\circ(y) + \alpha\|y\| \le 1\,\}$ is compact, the subdifferential formula follows immediately from Hiriart-Urruty and Lemaréchal [10, Theorem 4.4.2, Page 189]. The claim concerning $\kappa_\alpha(x)$ now follows from this and (4.2).

(ii) Suppose to the contrary that $\|x - \bar{x}\| < \alpha\kappa(\bar{x})$. Then $\kappa(\bar{x}) = \kappa_\alpha(x)$ and there exists $t \in (0, 1)$ so that

$$(1/\alpha)\|x - t\bar{x}\| < \kappa(t\bar{x}) < \kappa(\bar{x}) = \kappa_\alpha(x).$$

Hence, $\kappa_\alpha(x) > \max\{\,\kappa(t\bar{x}),\ (1/\alpha)\|x - t\bar{x}\|\,\} \ge \kappa_\alpha(x)$, which leads to a contradiction. Next, suppose also that κ is continuous and suppose to the contrary that $\|x - \bar{x}\| > \alpha\kappa(\bar{x})$. Then $(1/\alpha)\|x - \bar{x}\| = \kappa_\alpha(x)$ and there exists $t \in (0, 1)$ so that

$$\kappa_\alpha(x) = \alpha^{-1}\|x - \bar{x}\| > \alpha^{-1}(1 - t)\|x - \bar{x}\| = \alpha^{-1}\|x - (\bar{x} + t[x - \bar{x}])\| > \kappa(\bar{x} + t(x - \bar{x})).$$

Hence, $\kappa_\alpha(x) > \max\{\kappa(\bar{x}+t[x-\bar{x}]),\,(1/\alpha)\|x-(\bar{x}+t[x-\bar{x}])\|\} \geq \kappa_\alpha(x)$, which leads to a contradiction. This proves (ii).

(iii) Since κ is closed, the function $z \mapsto \max\{\,\kappa(z),\,(1/\alpha)\|x-z\|\,\}$ is closed and coercive for all x. Thus, $\mathsf{pprox}_{\alpha\kappa}(x)$ is nonempty for all x. Now, suppose that $\bar{x}$ and $\widehat{x}$ belong to $\mathsf{pprox}_{\alpha\kappa}(x)$. Since the function $z \mapsto \max\{\,\kappa(z),\,(1/\alpha)\|x-z\|\,\}$ is convex, it follows that $\{\,\lambda\widehat{x}+(1-\lambda)\bar{x} \mid \lambda\in[0,1]\,\} \subseteq \mathsf{pprox}_{\alpha\kappa}(x)$. Then, from part (ii),

$$\alpha\kappa_\alpha(x) = \|\widehat{x}-x\| = \|\bar{x}-x\| = \|\lambda\widehat{x}+(1-\lambda)\bar{x}-x\| \quad \forall\lambda\in(0,1). \tag{4.3}$$

Thus,

$$\alpha\kappa_\alpha(x) = \|\lambda\widehat{x}+(1-\lambda)\bar{x}-x\| \leq \lambda\|\widehat{x}-x\|+(1-\lambda)\|\bar{x}-x\| = \alpha\kappa_\alpha(x).$$

Hence, equality holds throughout. This implies that $\bar{x}-x$ and $\widehat{x}-x$ differ by a nonnegative scaling, and thus we may assume without loss of generality that $\bar{x}-x = \tau(\widehat{x}-x)$ for some $\tau\geq 0$. Now, if $\kappa_\alpha(x)>0$, we see from (4.3) that $\tau=1$, which implies $\bar{x}=\widehat{x}$. On the other hand, if $\kappa_\alpha(x)=0$, then (4.3) gives $\bar{x}=x=\widehat{x}$. Thus, $\mathsf{pprox}_{\alpha\kappa}(x)$ is a singleton.

Let $\gamma>0$. Then, for any x,

$$\begin{aligned}\mathsf{pprox}_{\alpha\kappa}(\gamma x) &= \arg\min_u \max\{\kappa(u),\,(1/\alpha)\|u-\gamma x\|\}\\ &= \gamma\arg\min_v \max\{\kappa(\gamma v),\,(1/\alpha)\|\gamma v-\gamma x\|\} \quad (u=\gamma v)\\ &= \gamma\arg\min_v \max\{\kappa(v),\,(1/\alpha)\|v-x\|\} = \gamma\,\mathsf{pprox}_{\alpha\kappa}(x),\end{aligned}$$

where the second to last equality holds because κ is positively homogeneous. Moreover, it is clear that $\mathsf{pprox}_{\alpha\kappa}(0)=0$. Thus, $\mathsf{pprox}_{\alpha\kappa}$ is positively homogeneous.

We next prove continuity. Let $x_k\to x$ and write $\bar{x}_k = \mathsf{pprox}_{\alpha\kappa}(x_k)$ for notational simplicity. Then we have from the definition of $\bar{x}_k$ that, for any u,

$$\max\{\kappa(\bar{x}_k),\,(1/\alpha)\|x_k-\bar{x}_k\|\} \leq \max\{\kappa(u),\,(1/\alpha)\|x_k-u\|\}. \tag{4.4}$$

By setting $u=0$ in (4.4), we immediately conclude that $\{\bar{x}_k\}$ is bounded. Take any convergent subsequence $\{\bar{x}_{k_i}\}$ of $\{\bar{x}_k\}$ and let $\widehat{x}$ denote its limit. Passing to the limit in (4.4) along the subsequences $\{x_{k_i}\}$ and $\{\bar{x}_{k_i}\}$ and invoking the closedness of κ gives

$$\max\{\kappa(\widehat{x}),\,(1/\alpha)\|x-\widehat{x}\|\} \leq \max\{\kappa(u),\,(1/\alpha)\|x-u\|\}$$

for any u. Thus, $\widehat{x} = \mathsf{pprox}_{\alpha\kappa}(x)$. Since any convergent subsequence of the bounded sequence $\{\,\mathsf{pprox}_{\alpha\kappa}(x_k)\,\}$ converges to $\mathsf{pprox}_{\alpha\kappa}(x)$, we have $\mathsf{pprox}_{\alpha\kappa}(x_k)\to\mathsf{pprox}_{\alpha\kappa}(x)$. This proves the continuity of $\mathsf{pprox}_{\alpha\kappa}$.

(iv) Consider any x satisfying $\kappa_\alpha(x)>0$. Recall from part (ii) that $\|x-\bar{x}\|\geq\alpha\kappa(\bar{x})$, where $\bar{x}=\mathsf{pprox}_{\alpha\kappa}(x)$ exists due to part (iii) and κ being closed. In particular, this implies

$$(1/\alpha)\|x-\bar{x}\| = \kappa_\alpha(x) > 0.$$

Combining this with Proposition 2.2, we see further that

$$\begin{aligned}\partial\kappa_\alpha(x) &= \bigcup_{\mu\in[0,1]}\left\{\partial(\mu^+\kappa)(\bar{x}) \cap \frac{1-\mu}{\alpha}\left\{\frac{x-\bar{x}}{\|x-\bar{x}\|}\right\} \;\middle|\; \mu^+\kappa(\bar{x}) + \frac{1-\mu}{\alpha}\|x-\bar{x}\| = \kappa_\alpha(x)\right\}\\ &\subseteq \left\{\lambda\cdot\frac{x-\bar{x}}{\|x-\bar{x}\|} \;\middle|\; \lambda\in[0,1/\alpha]\right\}.\end{aligned}$$

Thus, for any elements u and v in $\partial\kappa_\alpha(x)$, there exist scalars $\lambda_1, \lambda_2 \ge 0$ so that $u = \lambda_1(x - \bar{x})$ and $v = \lambda_2(x - \bar{x})$. In view of part (i) of this theorem,

$$0 < \kappa_\alpha(x) = \langle x, \lambda_1(x - \bar{x})\rangle = \langle x, \lambda_2(x - \bar{x})\rangle,$$

which establishes that $\langle x, x - \bar{x}\rangle > 0$ and $\lambda_1 = \lambda_2$. Hence, $\partial\kappa_\alpha(x)$ is a singleton whenever $\kappa_\alpha(x) > 0$, which implies that κ_α is differentiable at those x.

To obtain the formula for $\nabla\kappa_\alpha(x)$, note that $\nabla\kappa_\alpha(x) = \lambda(x - \bar{x})$ for some $\lambda \ge 0$. In view of part (i), we must have $\langle x, \lambda(x - \bar{x})\rangle = \kappa_\alpha(x) > 0$, and thus

$$\lambda = \frac{\kappa_\alpha(x)}{\langle x, x - \bar{x}\rangle} = \frac{\|x - \bar{x}\|}{\alpha\langle x, x - \bar{x}\rangle},$$

where the second equality is due to part (ii). This proves the formula for $\nabla\kappa_\alpha(x)$. ◻

5. Computing polar envelopes: Examples. Let κ be a closed gauge. We illustrate how to use Theorem 4.4 to compute $\mathsf{pprox}_{\alpha\kappa}(x)$—and hence $\kappa_\alpha(x)$—at those x with $\kappa_\alpha(x) > 0$, which are points of differentiability.

Recall from Theorem 4.4(iii) that the set $\mathsf{pprox}_{\alpha\kappa}(x)$ is a singleton. For any x that satisfies $\kappa_\alpha(x) > 0$, let $\bar{y} := \mathsf{pprox}_{\alpha\kappa}(x)$ and write $\bar{r} := \kappa_\alpha(x) > 0$. Then $\kappa(\bar{y}) \le (1/\alpha)\|\bar{y} - x\| = \bar{r}$ according to Theorem 4.4(ii). In particular, we have $\bar{y} \ne x$.

We consider two cases. First, suppose that $\kappa(\bar{y}) < (1/\alpha)\|\bar{y} - x\|$. It follows from Zălinescu [19, Corollary 2.8.15] that the optimality conditions for (4.1) are given by

$$0 \in \frac{1}{\alpha} \cdot \frac{\bar{y} - x}{\|\bar{y} - x\|} + \mathcal{N}(\bar{y} \mid \operatorname{dom}\kappa).$$

This implies $\bar{y} = \mathsf{proj}_{\operatorname{dom}\kappa}(x)$ and, in particular, that the projection exists in this case.

On the other hand, suppose that

$$\kappa(\bar{y}) = (1/\alpha)\|\bar{y} - x\| = \bar{r}. \tag{5.1}$$

Again applying Zălinescu [19, Corollary 2.8.15] to obtain the optimality condition for (4.1), we deduce that there exists $\lambda \in [0, 1]$ with the property that

$$0 \in \frac{1 - \lambda}{\alpha} \cdot \frac{\bar{y} - x}{\|\bar{y} - x\|} + \partial(\lambda^+\kappa)(\bar{y}). \tag{5.2}$$

We claim that $\lambda \ne 1$. Suppose to the contrary that $\lambda = 1$. Then (5.2) becomes $0 \in \partial\kappa(\bar{y})$. This means that $\bar{y}$ minimizes the gauge function κ, giving $\kappa(\bar{y}) = 0$. This contradicts (5.1) because $\bar{r} > 0$.

Thus, $\lambda \ne 1$ and we obtain from (5.2) that

$$0 \in \bar{y} - x + \frac{\alpha}{1 - \lambda}\|\bar{y} - x\|\partial(\lambda^+\kappa)(\bar{y}) \subset \bar{y} - x + \mathcal{N}(\bar{y} \mid [\kappa \le \bar{r}]),$$

where the second inclusion follows from Zălinescu [19, Corollary 2.9.5], (5.1), and the fact that $\bar{r} > 0$. This implies $\bar{y} = \mathsf{proj}_{[\kappa \le \bar{r}]}(x)$. Substituting this relation into (5.1) shows that $\bar{r}$ satisfies the equation

$$\alpha^2\bar{r}^2 = \|x - \mathsf{proj}_{[\kappa \le \bar{r}]}(x)\|^2. \tag{5.3}$$

Next, note that the function $r \mapsto \|x - \mathsf{proj}_{[\kappa\le r]}(x)\|^2$ is nonincreasing on $(0,\infty)$: indeed, for $s \ge r > 0$, we have

$$\begin{aligned}\|x - \mathsf{proj}_{[\kappa\le r]}(x)\|^2 &= \|x - \mathsf{proj}_{[\kappa\le s]}(x) + \mathsf{proj}_{[\kappa\le s]}(x) - \mathsf{proj}_{[\kappa\le r]}(x)\|^2\\ &= \|x - \mathsf{proj}_{[\kappa\le s]}(x)\|^2 + 2\langle x - \mathsf{proj}_{[\kappa\le s]}(x),\, \mathsf{proj}_{[\kappa\le s]}(x) - \mathsf{proj}_{[\kappa\le r]}(x)\rangle\\ &\quad + \|\mathsf{proj}_{[\kappa\le s]}(x) - \mathsf{proj}_{[\kappa\le r]}(x)\|^2\\ &\overset{\text{(a)}}{\ge} \|x - \mathsf{proj}_{[\kappa\le s]}(x)\|^2 + \|\mathsf{proj}_{[\kappa\le s]}(x) - \mathsf{proj}_{[\kappa\le r]}(x)\|^2 \ge \|x - \mathsf{proj}_{[\kappa\le s]}(x)\|^2,\end{aligned}$$

where the inequality (a) follows from the property of projections and the fact that $\mathsf{proj}_{[\kappa\le r]}(x) \in [\kappa \le s]$. Consequently, the function

$$r \mapsto \alpha^2 r^2 - \|x - \mathsf{proj}_{[\kappa\le r]}(x)\|^2$$

is *strictly* increasing on $(0,\infty)$. Thus, $\bar r$ is the *unique positive* root satisfying (5.3).

In summary,

$$\mathsf{pprox}_{\alpha\kappa}(x) = \begin{cases}\mathsf{proj}_{\mathrm{dom}\,\kappa}(x) & \text{if } \bar y := \mathsf{proj}_{\mathrm{dom}\,\kappa}(x) \text{ exists and } \kappa(\bar y) < (1/\alpha)\|x - \bar y\|,\\ \mathsf{proj}_{[\kappa\le \bar r]}(x) & \text{otherwise,}\end{cases}$$

$$\kappa_\alpha(x) = \begin{cases}(1/\alpha)\|x - \bar y\| & \text{if } \bar y := \mathsf{proj}_{\mathrm{dom}\,\kappa}(x) \text{ exists and } \kappa(\bar y) < (1/\alpha)\|x - \bar y\|,\\ \bar r & \text{otherwise,}\end{cases}$$

where $\bar r$ is the unique positive root of (5.3).

We now give examples that show how these formulas specialize to common cases.

Example 5.1 (linear function over a cone). Let $\kappa(z) = \langle c, z\rangle + \delta_{\mathcal K}(z)$ for some closed convex cone $\mathcal K$ and some vector c in the dual cone $\mathcal K^*$. Then κ is a closed gauge, $\mathrm{dom}\,\kappa = \mathcal K$, and $\mathsf{proj}_{\mathcal K}(x)$ exists for all x because $\mathcal K$ is closed. For any x satisfying $\kappa_\alpha(x) > 0$, the polar proximal map is given by

$$\mathsf{pprox}_{\alpha\kappa}(x) = \begin{cases}\mathsf{proj}_{\mathcal K}(x) & \text{if } \langle c, \mathsf{proj}_{\mathcal K}(x)\rangle < (1/\alpha)\|\mathsf{proj}_{\mathcal K^\circ}(x)\|,\\ \mathsf{proj}_{\overline{\mathcal K}(\bar r)}(x) & \text{otherwise,}\end{cases}$$

$$\kappa_\alpha(x) = \begin{cases}(1/\alpha)\|\mathsf{proj}_{\mathcal K^\circ}(x)\| & \text{if } \langle c, \mathsf{proj}_{\mathcal K}(x)\rangle < (1/\alpha)\|\mathsf{proj}_{\mathcal K^\circ}(x)\|,\\ \bar r & \text{otherwise,}\end{cases}$$

where $\overline{\mathcal K}(\bar r) := \mathcal K \cap \{\, u \mid \langle c, u\rangle \le \bar r \,\}$ and $\bar r$ is the unique positive root of the equation

$$\alpha^2 \bar r^2 = \|x - \mathsf{proj}_{\overline{\mathcal K}(\bar r)}(x)\|^2.$$

Here we used the Moreau identity to determine that $x = \mathsf{proj}_{\mathcal K}(x) + \mathsf{proj}_{\mathcal K^\circ}(x)$. ◻

Example 5.2 (continuous gauge). Let κ be a continuous gauge. Then we have from Theorem 4.4(ii) that $\kappa(\mathsf{pprox}_{\alpha\kappa}(x)) = (1/\alpha)\|\mathsf{pprox}_{\alpha\kappa}(x) - x\|$. Then, for any x satisfying $\kappa_\alpha(x) > 0$, it holds that

$$\kappa_\alpha(x) = \bar r \quad \text{and} \quad \mathsf{pprox}_{\alpha\kappa}(x) = \mathsf{proj}_{[\kappa\le \bar r]}(x),$$

where $\bar r$ is the unique positive root of the equation

$$\alpha^2 \bar r^2 = \|x - \mathsf{proj}_{[\kappa\le \bar r]}(x)\|^2. \tag{5.4}$$

◻

Example 5.3 (infinity norm). As a concrete example, consider $\kappa = \|\cdot\|_\infty$. Note that $\kappa_\alpha(x) = 0$ if and only if $x = 0$. For $x \neq 0$, (5.4) becomes

$$\alpha^2 \bar{r}^2 = \sum_{i=1}^{n} (|x_i| - \bar{r})_+^2.$$

Letting $\bar{\gamma} = \bar{r}^{-1}$, the above equation is equivalent to

$$\alpha^2 = \sum_{i=1}^{n} (\bar{\gamma}|x_i| - 1)_+^2.$$

This is a piecewise linear quadratic equation with exactly one positive root because the function on the right-hand side is zero for $\bar{\gamma} \in (0, 1/\|x\|_\infty]$ and is strictly increasing on $(1/\|x\|_\infty, \infty)$, mapping this interval to $(0, \infty)$. The equation can be solved in $O(n \log n)$ time. Once $\bar{r}$ is found, we can obtain $\mathsf{pprox}_{\alpha\kappa}(x) = \mathsf{proj}_{[\|\cdot\|_\infty \leq \bar{r}]}(x)$. ☐

6. Constructing smooth gauge dual problems. The polar envelope can be naturally incorporated in gauge optimization problems to yield smooth gauge dual problems. We show how a primal solution can be recovered after solving the smooth dual problem. First, we collect in the following proposition some essential properties of the polar proximal map of a continuous gauge.

PROPOSITION 6.1 (polar proximal map of a continuous gauge). *Let κ be a continuous gauge and $\alpha > 0$. Then*

$$\mathsf{pprox}_{\alpha\kappa}(x) = \mathsf{proj}_{[\kappa \leq \kappa_\alpha(x)]}(x). \tag{6.1}$$

Moreover, the following hold.

(i) $\|\mathsf{pprox}_{\alpha\kappa}(x)\| \leq \|x\|$ *for all x.*

(ii) *For any $\beta > 0$ and $M > 0$, the function $x \mapsto \mathsf{pprox}_{\alpha\kappa}(x)$ is globally Lipschitz continuous in the set $\Xi_{M,\beta} := \{\, x \mid \|x\| \leq M, \kappa_\alpha(x) \geq \beta \,\}$. Specifically, it holds that*

$$\|\mathsf{pprox}_{\alpha\kappa}(x) - \mathsf{pprox}_{\alpha\kappa}(y)\| \leq \frac{3M}{\alpha\beta}\|x - y\|$$

for any x and y in $\Xi_{M,\beta}$.

(iii) *At any x satisfying $\kappa_\alpha(x) > 0$, the function $\nabla \kappa_\alpha$ is locally Lipschitz.*

Proof. We first prove (6.1). In view of Example 5.2, it suffices to show that (6.1) also holds in the case when $\kappa_\alpha(x) = 0$. Fix any such x. Write $\bar{y} = \mathsf{pprox}_{\alpha\kappa}(x)$. Then we have $\kappa(\bar{y}) = 0$ and $\|\bar{y} - x\| = 0$. Consequently, we have $\bar{y} = x$ and that $\kappa(x) = 0$. In particular, this indicates that we can write $\bar{y} = x = \mathsf{proj}_{[\kappa \leq 0]}(x) = \mathsf{proj}_{[\kappa \leq \kappa_\alpha(x)]}(x)$. This proves (6.1).

We now prove part (i). In view of (6.1) and the fact that $\kappa(0) \leq \kappa_\alpha(x)$ for any x, we have from the definition of projection that

$$\langle x - \mathsf{pprox}_{\alpha\kappa}(x),\, 0 - \mathsf{pprox}_{\alpha\kappa}(x)\rangle \leq 0,$$

which implies that $\|\mathsf{pprox}_{\alpha\kappa}(x)\|^2 \leq \langle x, \mathsf{pprox}_{\alpha\kappa}(x)\rangle$. The conclusion of part (i) now follows from this and the Cauchy–Schwarz inequality.

Next, in view of (6.1) and the nonexpansiveness of projections onto closed convex sets, it follows that

$$\|\mathsf{pprox}_{\alpha\kappa}(x) - \mathsf{pprox}_{\alpha\kappa}(y)\| \leq \|x - y\| \tag{6.2}$$

whenever $\kappa_\alpha(x) = \kappa_\alpha(y)$. Now, consider any $x,\, y \in \Xi_{M,\beta}$. Thus,

$$\begin{aligned}
&\|\mathsf{pprox}_{\alpha\kappa}(x) - \mathsf{pprox}_{\alpha\kappa}(y)\| \\
&\le \|\mathsf{pprox}_{\alpha\kappa}(x) - \mathsf{pprox}_{\alpha\kappa}(\kappa_\alpha(x)y/\kappa_\alpha(y))\| + \|\mathsf{pprox}_{\alpha\kappa}(\kappa_\alpha(x)y/\kappa_\alpha(y)) - \mathsf{pprox}_{\alpha\kappa}(y)\| \\
&= \|\mathsf{pprox}_{\alpha\kappa}(x) - \mathsf{pprox}_{\alpha\kappa}(\kappa_\alpha(x)y/\kappa_\alpha(y))\| + \|\kappa_\alpha(x)\,\mathsf{pprox}_{\alpha\kappa}(y/\kappa_\alpha(y)) - \mathsf{pprox}_{\alpha\kappa}(y)\| \\
&\le \|x - \kappa_\alpha(x)y/\kappa_\alpha(y)\| + \|\kappa_\alpha(x)\,\mathsf{pprox}_{\alpha\kappa}(y/\kappa_\alpha(y)) - \mathsf{pprox}_{\alpha\kappa}(y)\| \\
&= \kappa_\alpha(y)^{-1}\big[\,\|\kappa_\alpha(y)x - \kappa_\alpha(x)y\| + \|\mathsf{pprox}_{\alpha\kappa}(y)\|\cdot|\kappa_\alpha(y) - \kappa_\alpha(x)|\,\big],
\end{aligned}$$

where the equalities follow from the positive homogeneity of $\mathsf{pprox}_{\alpha\kappa}$ (cf. Theorem 4.4(iii)), and the second inequality follows from (6.2). Next, using part (i),

$$\begin{aligned}
\|\mathsf{pprox}_{\alpha\kappa}(x) - \mathsf{pprox}_{\alpha\kappa}(y)\| &\le \kappa_\alpha(y)^{-1}\big[\,\|\kappa_\alpha(y)x - \kappa_\alpha(x)y\| + \|y\|\cdot|\kappa_\alpha(y) - \kappa_\alpha(x)|\,\big] \\
&\overset{\text{(a)}}{\le} \kappa_\alpha(y)^{-1}\big[\,\|\kappa_\alpha(y)(x-y)\| + 2\|y\|\cdot|\kappa_\alpha(x) - \kappa_\alpha(y)|\,\big] \\
&\overset{\text{(b)}}{\le} \beta^{-1}\big[\,\|\kappa_\alpha(y)(x-y)\| + 2M|\kappa_\alpha(x) - \kappa_\alpha(y)|\,\big] \\
&\overset{\text{(c)}}{\le} \frac{3M}{\alpha\beta}\|x-y\|,
\end{aligned}$$

where (a) follows from the triangle inequality, (b) follows from the definition of $\Xi_{M,\beta}$, and (c) follows from the Lipschitz continuity of κ_α (cf. Proposition 4.3(i)). This proves global Lipschitz continuity on $\Xi_{M,\beta}$. Finally, the conclusion of part (iii) follows immediately from this and the formula for $\nabla\kappa_\alpha$ given in Theorem 4.4(iv). ∎

6.1. Sublinear regularization. Proposition 6.1 suggests a natural smoothing strategy for the following gauge optimization problem

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad \rho(b - Ax) \le \sigma, \tag{6.3a}$$

where κ and ρ are both closed gauges and $A : \mathcal{X} \to \mathcal{Y}$ is a linear map. We assume that $\sigma \in [0, \rho(b))$, that $\kappa^{-1}(0) = \{\,0\,\}$ and $\rho^{-1}(0) = \{\,0\,\}$, and that the data satisfy the constraint qualification

$$\text{ri dom}\,\kappa \cap A^{-1}\text{ri}\,\mathcal{C} \ne \emptyset. \tag{6.3b}$$

Here, we define $\mathcal{C} := \{\, u \mid \rho(b - u) \le \sigma \,\}$.

Consider the following regularization of (6.3a):

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \kappa(x) + \alpha\|x\| \quad \text{subject to} \quad \rho(b - Ax) \le \sigma, \tag{6.4}$$

where α is a positive regularization parameter. We expect that, for small values of α, solutions of the perturbed problem are good approximations to solutions of the original problem (6.3). In particular, because the objective $\kappa + \alpha\|\cdot\|$ epi-converges to κ [15, Proposition 7.4(c)], it follows that cluster points (if they exist) of solutions of (6.4) are also minimizers of (6.3); see [15, Theorem 7.31(b)].

The objective of the perturbed problem (6.4) is in general a nonsmooth gauge. As we demonstrate below, however, the gauge dual has a smooth objective.

6.2. Primal and dual pairs under convolution. The corresponding gauge dual is given by

$$\underset{y\in\mathcal{Y}}{\text{minimize}} \quad (\kappa^\circ \diamond (1/\alpha)\|\cdot\|)(A^*y) \quad \text{subject to} \quad \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1, \tag{6.5}$$

where the objective in (6.5) follows from Friedlander, Macêdo, and Pong [9, section 1.1] and an application of Lemma 3.3 with $\kappa_1 = \kappa^\circ$ and $\kappa_2 = (1/\alpha)\|\cdot\|$. The next result concerning the gauge dual pair (6.4) and (6.5) establishes that the dual problem is in some sense smooth, and gives a formula for the relationship between the primal and dual solutions.

THEOREM 6.2 (primal and dual problems under polar convolution). *Consider problems* (6.4) *and* (6.5)*, where* κ, ρ, A, b, *and* σ *are as in* (6.3)*. Then the following conclusions hold.*

(i) *The optimal values of* (6.4) *and* (6.5) *are finite, positive, and attained. Moreover, they are the reciprocal of one another.*
(ii) *The objective of* (6.5) *is smooth with a locally Lipschitz gradient on the feasible set of* (6.5)*.*
(iii) *Let* $\bar{y}$ *be an optimal solution of* (6.5) *and* $\bar{r}$ *be its optimal value. Then*

$$\widehat{x} = \frac{1/\bar{r}}{\kappa(\mathsf{prox}_{\bar{r}\kappa}(A^*\bar{y})) + \alpha\|\mathsf{prox}_{\bar{r}\kappa}(A^*\bar{y})\|}\,\mathsf{prox}_{\bar{r}\kappa}(A^*\bar{y}) \tag{6.6}$$

is an optimal solution of (6.4)*.*

Proof. Note from (6.3b) that we have

$$\operatorname{ri}\operatorname{dom}(\kappa + \alpha\|\cdot\|) \cap A^{-1}\operatorname{ri}\mathcal{C} = \operatorname{ri}\operatorname{dom}\kappa \cap A^{-1}\operatorname{ri}\mathcal{C} \neq \emptyset.$$

Also, from Proposition 4.3(i), we trivially have $\operatorname{ri}\operatorname{dom}(\kappa^\circ \diamond (1/\alpha)\|\cdot\|) \cap A^*\operatorname{ri}\mathcal{C}' = A^*\operatorname{ri}\mathcal{C}' \neq \emptyset$, where $\mathcal{C}'$ is the antipolar set of $\mathcal{C}$, which is a nonempty set because $0 \notin \mathcal{C}$, thanks to the assumption that $\rho(b) > \sigma$. In view of these and Friedlander, Macêdo, and Pong [9, Corollary 5.6], the optimal value of (6.5) is the reciprocal of the optimal value of (6.4), and both optimal values are finite, positive, and attained.

Next, since $\kappa^{-1}(0) = \{0\}$, it follows that κ° is continuous. Moreover, because the optimal value of (6.5) is attained and is positive, it follows that $(\kappa^\circ \diamond (1/\alpha)\|\cdot\|)(A^*y) > 0$ for any y feasible for (6.5). The local Lipschitz continuity of the gradient of the objective of (6.5) now follows from this and Proposition 6.1(iii).

We now prove part (iii). Let $\bar{y}$ be an optimal solution of (6.5) and let $\bar{u}$ satisfy

$$\kappa^\circ(\bar{u}) = (1/\alpha)\|A^*\bar{y} - \bar{u}\| = (\kappa^\circ \diamond (1/\alpha)\|\cdot\|)(A^*\bar{y}) =: \bar{r} > 0, \tag{6.7}$$

which exists according to Theorem 4.4(ii) and (iii) and the continuity of κ°. Let $\widehat{x}$ be a solution of (6.4) and let $\widehat{v}$ be such that $b = A\widehat{x} + \widehat{v}$ and $\rho(\widehat{v}) \le \sigma$. Then we have from strong duality of the gauge dual pairs (6.4) and (6.5) that

$$\begin{aligned} 1 &= (\kappa(\widehat{x}) + \alpha\|\widehat{x}\|) \cdot (\kappa^\circ \diamond (1/\alpha)\|\cdot\|)(A^*\bar{y}) \\ &\overset{(a)}{=} \kappa(\widehat{x}) \cdot \kappa^\circ(\bar{u}) + \|\widehat{x}\| \cdot \|A^*\bar{y} - \bar{u}\| \\ &\overset{(b)}{\ge} \langle \widehat{x}, \bar{u}\rangle + \langle \widehat{x}, A^*\bar{y} - \bar{u}\rangle = \langle \widehat{x}, A^*\bar{y}\rangle \\ &\overset{(c)}{=} \langle b - \widehat{v}, \bar{y}\rangle \overset{(d)}{\ge} \langle b, \bar{y}\rangle - \sigma\rho^\circ(\bar{y}) \ge 1, \end{aligned} \tag{6.8}$$

where (a) follows from (6.7), (b) follows from (3.2), (c) follows from the fact that $A\widehat{x} + \widehat{v} = b$, and (d) follows from Friedlander, Macêdo, and Pong [9, Proposition 2.1(iii)] (for $\sigma > 0$) and the continuity of ρ° (for $\sigma = 0$), thanks to the assumption that $\rho^{-1}(0) = \{0\}$. Thus, equality holds throughout the above relation, and we have in particular that

$$\|\widehat{x}\| \cdot \|A^*\bar{y} - \bar{u}\| = \langle \widehat{x}, A^*\bar{y} - \bar{u}\rangle.$$

Since $\widehat{x} \neq 0$, this implies that $\widehat{x} = \gamma(A^*\bar{y} - \bar{u})$ for some $\gamma > 0$. Next, combine (6.7) with the first equation in (6.8) to obtain $\kappa(\widehat{x}) + \alpha\|\widehat{x}\| = \bar{r}^{-1}$. Together with the expression that we have just derived for $\widehat{x}$, we deduce that

$$\gamma = \frac{1/\bar{r}}{\kappa(A^*\bar{y} - \bar{u}) + \alpha\|A^*\bar{y} - \bar{u}\|}.$$

Recall from (6.1) that $\bar{u} = \mathsf{proj}_{\kappa^\circ(\cdot)\leq\bar{r}}(A^*\bar{y})$ because κ° is continuous. Thus, by the Moreau identity, $A^*\bar{y} - \bar{u} = \mathsf{prox}_{\bar{r}\kappa}(A^*\bar{y})$. Then we can compute $\widehat{x}$ as in (6.6). ∎

As an immediate application of Theorem 6.2, we consider the basis pursuit problem [7], which takes the form

$$\text{(6.9)} \qquad \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \|x\|_1 \quad \text{subject to} \quad Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$, $A^{-1}\{\,b\,\} \neq \emptyset$, and $b \neq 0$. This is just (6.3a) with $\kappa = \|\cdot\|_1$, $\rho = \delta_{\{0\}}$, and $\sigma = 0$, and it is routine to check that the assumptions on κ, ρ, A, b, and σ for (6.3) are satisfied. For each $\alpha > 0$, one can consider the regularization of (6.9),

$$\text{(6.10)} \qquad \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \|x\|_1 + \alpha\|x\| \quad \text{subject to} \quad Ax = b,$$

and its gauge dual problem

$$\text{(6.11)} \qquad \underset{y\in\mathbb{R}^m}{\text{minimize}} \quad (\|\cdot\|_\infty \diamond (1/\alpha)\|\cdot\|)(A^*y) \quad \text{subject to} \quad \langle b, y\rangle \geq 1.$$

We thus have the following immediate corollary of Theorem 6.2.

Corollary 6.3 (polar smoothing of the gauge dual). *Consider the gauge dual pair* (6.10) *and* (6.11) *with $A^{-1}\{\,b\,\} \neq \emptyset$ and $b \neq 0$. Then the following conclusions hold.*

- (i) *The optimal values of* (6.10) *and* (6.11) *are finite, positive, and attained. Moreover, they are the reciprocal of one other.*
- (ii) *The objective of* (6.11) *is smooth with a locally Lipschitz gradient on the feasible set of* (6.11).
- (iii) *Let $\bar{y}$ be an optimal solution of* (6.11) *and $\bar{r}$ be its optimal value. Then*

$$\text{(6.12)} \qquad \widehat{x} = \frac{1/\bar{r}}{\|\,\mathsf{prox}_{\bar{r}\|\cdot\|_1}(A^*\bar{y})\|_1 + \alpha\|\,\mathsf{prox}_{\bar{r}\|\cdot\|_1}(A^*\bar{y})\|}\,\mathsf{prox}_{\bar{r}\|\cdot\|_1}(A^*\bar{y})$$

is an optimal solution of (6.10).

Corollary 6.3(ii) suggests that, in order to solve (6.10), one can apply a gradient descent algorithm with line search to solve (6.11), which is appropriate because the gradient is locally Lipschitz. A solution of (6.10) is then recovered via (6.12).

As we discussed in connection with the duality correspondences shown in Figures 1.1 and 1.2, we contrast the above approach with the usual way of constructing smooth *Lagrange* dual problems by adding a *strongly convex* term to the primal objective. In this latter approach, for the functions and data in (6.3) and $\alpha > 0$, one considers the following approximation of (6.3a):

$$\text{(6.13)} \qquad \underset{x\in\mathcal{X}}{\text{minimize}} \quad \kappa(x) + (\alpha/2)\|x\|^2 \quad \text{subject to} \quad \rho(b - Ax) \leq \sigma.$$

The Lagrange dual problem of (6.13) takes the form

$$\text{(6.14)} \qquad \underset{y\in\mathcal{Y}}{\text{maximize}} \quad \langle b, y\rangle - (1/2\alpha)\,\mathsf{dist}^2_{[\kappa^\circ\leq 1]}(A^*y) - \sigma\rho^\circ(y).$$

This dual problem is the sum of the nonsmooth function $y \mapsto -\sigma\rho^\circ(y)$ and a smooth function with Lipschitz gradient, and thus can be suitably solved by first-order methods such as the proximal-gradient algorithm. The solution $\bar{x}$ of (6.13) is then recovered from a solution $\bar{y}$ of (6.14) via

$$\bar{x} = \mathsf{prox}_{(1/\alpha)\kappa}(\alpha^{-1}A^*\bar{y}).$$

In contrast to the gauge dualization approach based on the pair (6.4) and (6.5), the Lagrange dual pair (6.13) and (6.14) are not gauge optimization problems, even though the original problem (6.3a) that we are approximating is a gauge optimization problem. Moreover, while the objective of (6.13) is strongly convex so that the objective of its Lagrange dual (6.14) has a smooth component, it is interesting to note that the objective of (6.4) is *not strongly convex* but its gauge dual (6.5) still has an objective that is smooth in the feasible region.

7. Proximal-point-like algorithms based on polar envelopes. In this section, we discuss proximal-point-like algorithms for solving the following general convex optimization problem:

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad f(x), \tag{7.1a}$$

where $f : \mathcal{X} \to \mathbb{R}_+ \cup \{+\infty\}$ is a proper closed nonnegative convex function with

$$\inf f > 0 \quad \text{and} \quad \arg\min f \neq \emptyset. \tag{7.1b}$$

Our proximal-point-like algorithms are based on a specific polar envelope and the corresponding polar proximal map. To proceed, we first rewrite (7.1a) as the equivalent gauge optimization problem

$$\underset{x\in\mathcal{X},\,\lambda=1}{\text{minimize}} \quad f^\pi(x,\lambda),$$

where the perspective function f^π is defined in (2.2).

7.1. The projected polar proximal map. We introduce two important ingredients for the development of our proximal-point-like algorithms, namely the *projected* polar envelope and the *projected* polar proximal map of $f^\pi(x,\lambda)$: for any $\alpha > 0$,

$$\begin{aligned}\mathsf{p}_{\alpha,f}(x) &:= f^\pi_\alpha(x,1),\\ \mathfrak{P}_{\alpha,\mathrm{f}}(x) &:= \mathsf{pprox}_{\alpha f^\pi}(x,1),\end{aligned}$$

where f^π_α is shorthand for the polar envelope of the perspective transform, i.e., $f^\pi_\alpha \equiv (f^\pi)_\alpha$. Note that $\mathfrak{P}_{\alpha,\mathrm{f}}(x)$ is a singleton for any x because αf^π is closed and by virtue of Theorem 4.4(iii). We show that the set of minimizers of the projected polar envelope and the set of "fixed points" of the polar proximal map are closely related to the set of minimizers of the original problem (7.1a).

We need the following auxiliary lemma concerning $\operatorname{dom} f^\pi$. For this section only, for any convex set $\mathcal{C}$ we define its negative polar cone by $\mathcal{C}^- = (\operatorname{cone}\mathcal{C})^\circ$.

LEMMA 7.1. *Consider f as in* (7.1). *Then it holds that*

$$\operatorname{cl}\operatorname{dom} f^\pi = \operatorname{cl}\left(\bigcup_{\lambda>0}\lambda(\operatorname{dom} f\times\{1\})\right). \tag{7.2}$$

Moreover, for any $(\bar{x},\bar{\lambda}) \in \operatorname{dom} f^\pi$, it holds that

$$\mathcal{N}((\bar{x},\bar{\lambda}) \mid \operatorname{dom} f^\pi) = \{\, w \in [\operatorname{dom} f\times\{1\}]^- \mid \langle w,(\bar{x},\bar{\lambda})\rangle = 0 \,\}. \tag{7.3}$$

Proof. First of all, we note from the definition of the recession cone that, for any closed convex set $\mathcal{C}$ and linear map A so that $A(\mathcal{C})$ is well defined, one has

$$A(\mathcal{C}_\infty) \subseteq [\operatorname{cl} A(\mathcal{C})]_\infty.$$

Applying this with $\mathcal{C} = \operatorname{epi} f$ and $A = \mathsf{proj}_{\mathcal{X}}$ and invoking $\operatorname{epi} f_\infty = (\operatorname{epi} f)_\infty$ gives

$$\operatorname{dom} f_\infty = \mathsf{proj}_{\mathcal{X}}(\operatorname{epi} f_\infty) \subseteq [\operatorname{cl} \mathsf{proj}_{\mathcal{X}}(\operatorname{epi} f)]_\infty = [\operatorname{cl} \operatorname{dom} f]_\infty. \tag{7.4}$$

Next, it is not hard to see directly from the definition of f^π that

$$\operatorname{dom} f^\pi = (\operatorname{dom} f_\infty \times \{0\}) \cup \bigcup_{\lambda > 0} \lambda(\operatorname{dom} f \times \{1\}). \tag{7.5}$$

Then we deduce further that

$$\begin{aligned} \operatorname{cl} \operatorname{dom} f^\pi \supseteq \operatorname{cl}\left(\bigcup_{\lambda>0} \lambda(\operatorname{dom} f \times \{1\})\right) &\overset{\text{(a)}}{=} \operatorname{cl}\left(\bigcup_{\lambda>0} \lambda(\operatorname{cl}\operatorname{dom} f \times \{1\})\right) \\ &\overset{\text{(b)}}{=} ([\operatorname{cl}\operatorname{dom} f]_\infty \times \{0\}) \cup \bigcup_{\lambda>0} \lambda(\operatorname{cl}\operatorname{dom} f \times \{1\}) \overset{\text{(c)}}{\supseteq} \operatorname{dom} f^\pi, \end{aligned}$$

where (a) follows directly from the definition of closure, (b) follows from Auslender and Teboulle [3, Lemma 2.1.1], and (c) follows from (7.4) and (7.5). Taking the closure on both sides of the above inclusion establishes (7.2).

Next, using (7.2) and the definition of negative polar, we have

$$[\operatorname{dom} f^\pi]^- = \left[\operatorname{cl}\left(\bigcup_{\lambda>0} \lambda(\operatorname{dom} f \times \{1\})\right)\right]^- = [\operatorname{dom} f \times \{1\}]^-.$$

The relation (7.3) now follows from this, the definition of normal cones, and the fact that $\operatorname{dom} f^\pi$ is a cone. ◻

In our next theorem, we study the relationship between the minimizers of the projected polar envelope $\mathsf{p}_{\alpha,f}$ and those of the original problem (7.1a). This result depends on the subdifferential of the perspective function, characterized below. We refer the reader to Combettes [8, Proposition 2.3(v)] and Aravkin et al. [1, Lemma 3.8] for a proof.

Proposition 7.2 (subdifferential of perspective). *Suppose that $f : \mathcal{X} \to \mathbb{R}_+ \cup \{+\infty\}$ is a proper closed nonnegative convex function. Then, for fixed $(x, \lambda) \in \operatorname{dom} f^\pi$,*

$$\partial f^\pi(x,\lambda) = \begin{cases} \{\, (u, -f^*(u)) \mid u \in \partial f(\lambda^{-1}x) \,\} & \text{if } \lambda > 0, \\ \{\, (u,\gamma) \mid (u,-\gamma) \in \operatorname{epi} f^*, u \in \partial f_\infty(x) \,\} & \text{if } \lambda = 0. \end{cases}$$

Theorem 7.3 (minimizers of the projected polar envelope). *Consider f as in (7.1) and let $\alpha > 0$.*

(i) *Suppose $x \in \operatorname{arg\,min} \mathsf{p}_{\alpha,f}$ and let $(\bar{x}, \bar{\lambda}) = \mathfrak{P}_{\alpha,\mathrm{f}}(x)$. Then $\bar{x} = x$, $\bar{\lambda} > 0$, $\bar{\lambda}^{-1}x \in \operatorname{arg\,min} f$.*

(ii) *If $x \in \operatorname{arg\,min} f$, then $\lambda x \in \operatorname{arg\,min} \mathsf{p}_{\alpha,f}$, where $\lambda := [1 + \alpha f(x)]^{-1}$.*

Proof. We first derive a formula for the subdifferential of $\mathsf{p}_{\alpha,f}$. To this end, define

$$f_1 = (1/\alpha)\|\cdot\| \quad \text{and} \quad f_2 = f^\pi. \tag{7.6}$$

Then $f_1 \diamond f_2$ is continuous thanks to Proposition 3.2, and $\mathsf{p}_{\alpha,f}(x) = (f_1 \diamond f_2)(x,1)$. We thus obtain the following formula for the subdifferential:

$$\partial \mathsf{p}_{\alpha,f}(x) = \{\, u \mid \exists \beta \ \text{ s.t. } \ (u,\beta) \in \partial(f_1 \diamond f_2)(x,1) \,\}\,; \tag{7.7}$$

here, $\partial(f_1 \diamond f_2)(x,1)$ is given by Proposition 2.2 as

$$\partial(f_1 \diamond f_2)(x,1) = \bigcup_{\mu \in \Gamma} \{\, \partial(\mu f_1)(x-\bar{x}, 1-\bar{\lambda}) \cap \partial([1-\mu]^+ f_2)(\bar{x},\bar{\lambda}) \,\}\,, \tag{7.8}$$

where $\Gamma := \{\, \mu \in [0,1] \mid \mu f_1(x-\bar{x}, 1-\bar{\lambda}) + (1-\mu)^+ f_2(\bar{x},\bar{\lambda}) = (f_1 \diamond f_2)(x,1) \,\}$.

We now prove part (i). Suppose that $x \in \arg\min \mathsf{p}_{\alpha,f}$ and let $(\bar{x},\bar{\lambda}) = \mathfrak{P}_{\alpha,\mathfrak{f}}(x)$. Then $0 \in \partial \mathsf{p}_{\alpha,f}(x)$. In view of Theorem 4.4(ii), there are two cases to consider, given below.

Case 1. $f^\pi(\bar{x},\bar{\lambda}) < \alpha^{-1}\|(x-\bar{x}, 1-\bar{\lambda})\|$. This forces $\mu = 1$ in (7.8), which together with $0 \in \partial \mathsf{p}_{\alpha,f}(x)$ and (7.7) implies the existence of β such that

$$(0,\beta) \in \alpha^{-1}\partial\|(x-\bar{x}, 1-\bar{\lambda})\| \cap \mathcal{N}((\bar{x},\bar{\lambda}) \mid \operatorname{dom} f^\pi).$$

In particular, we have $x = \bar{x}$. Since $\alpha^{-1}\|(x-\bar{x}, 1-\bar{\lambda})\| > 0$, we must then have $\bar{\lambda} \neq 1$. Thus,

$$\beta = \frac{1-\bar{\lambda}}{\alpha|1-\bar{\lambda}|}.$$

Combining this with (7.3) and $(0,\beta) \in \mathcal{N}((\bar{x},\bar{\lambda}) \mid \operatorname{dom} f^\pi)$ yields

$$0 \geq \big\langle (y,1), (0, 1-\bar{\lambda}) \big\rangle = 1-\bar{\lambda} \ \text{ and } \ \big\langle (\bar{x},\bar{\lambda}), (0, 1-\bar{\lambda}) \big\rangle = 0$$

for any fixed $y \in \operatorname{dom} f$. The first relation above together with $\bar{\lambda} \neq 1$ yields $\bar{\lambda} > 1$, and hence we conclude that $\bar{\lambda} = 0$ by the second relation, which is a contradiction. Consequently, this case cannot happen.

Case 2. $f^\pi(\bar{x},\bar{\lambda}) = \alpha^{-1}\|(x-\bar{x}, 1-\bar{\lambda})\|$. In this case, we see from $0 \in \partial \mathsf{p}_{\alpha,f}(x)$, (7.7), and (7.8) that there exist β and $\mu \in [0,1]$ satisfying

$$(0,\beta) \in (\mu/\alpha)\partial\|(x-\bar{x}, 1-\bar{\lambda})\| \cap \partial([1-\mu]^+ f^\pi)(\bar{x},\bar{\lambda}). \tag{7.9}$$

We claim that $\mu > 0$. Assume to the contrary that $\mu = 0$. It then follows from (7.9) that $\beta = 0$ and hence

$$0 \in \partial f^\pi(\bar{x},\bar{\lambda}),$$

meaning that $(\bar{x},\bar{\lambda})$ minimizes the gauge function f^π. Thus, $f^\pi(\bar{x},\bar{\lambda}) = 0$. Because $f^\pi(\bar{x},\bar{\lambda}) = \alpha^{-1}\|(x-\bar{x}, 1-\bar{\lambda})\|$, we see further that $(\bar{x},\bar{\lambda}) = (x,1)$. Hence,

$$0 = f^\pi(\bar{x},\bar{\lambda}) = f^\pi(x,1) = f(x).$$

Since the optimal value of (7.1a) is positive, we have arrived at a contradiction. Consequently, we have shown that $\mu > 0$.

From $\mu > 0$ and (7.9), we see that $(0,\beta) \in (\mu/\alpha)\partial\|(x-\bar{x}, 1-\bar{\lambda})\|$, which readily gives $x = \bar{x}$.

Next, we claim that $\bar{\lambda} \neq 1$. Indeed, suppose to the contrary that $\bar{\lambda} = 1$. Then

$$f(\bar{x}) = f^\pi(\bar{x},1) = f^\pi(\bar{x},\bar{\lambda}) = \alpha^{-1}\|(x-\bar{x}, 1-\bar{\lambda})\| = 0$$

since $x = \bar{x}$. Because the optimal value of (7.1a) is positive, this is a contradiction. Consequently, we must have $\bar{\lambda} \neq 1$.

Now, we claim that $\mu < 1$. Suppose to the contrary that $\mu = 1$. Then we have from (7.9) and (7.3) that

$$(0,\beta) \in \partial[0^+ f^\pi](\bar{x},\bar{\lambda}) = \mathcal{N}((\bar{x},\bar{\lambda}) \mid \operatorname{dom} f^\pi) \\ = \{\, w \in [\operatorname{dom} f \times \{1\}]^- \mid \langle w, (\bar{x},\bar{\lambda})\rangle = 0 \,\}. \tag{7.10}$$

In addition, we have from (7.9) that $(0,\beta) \in (1/\alpha)\partial\|(x-\bar{x}, 1-\bar{\lambda})\|$, which together with $x = \bar{x}$ and $\bar{\lambda} \neq 1$ gives

$$\beta = \frac{1-\bar{\lambda}}{\alpha|1-\bar{\lambda}|}. \tag{7.11}$$

Equations (7.10) and (7.11) together imply that

$$0 \geq \langle (y,1), (0, 1-\bar{\lambda})\rangle = 1-\bar{\lambda} \quad \text{and} \quad 0 = \langle (\bar{x},\bar{\lambda}), (0,1-\bar{\lambda})\rangle$$

for any $y \in \operatorname{dom} f$. As in case 1, this yields a contradiction. Thus, we must also have $\mu < 1$.

To summarize, we have shown that $\mu \in (0,1)$, $x = \bar{x}$, and $\bar{\lambda} \neq 1$. Together with (7.9), we conclude that

$$\frac{1}{1-\mu}(0,\beta) = \frac{\mu}{\alpha(1-\mu)}\left(0, \frac{1-\bar{\lambda}}{|1-\bar{\lambda}|}\right) \in \partial f^\pi(\bar{x},\bar{\lambda}). \tag{7.12}$$

We claim that $\bar{\lambda} > 0$. Suppose to the contrary that $\bar{\lambda} = 0$. Then, due to $f^\pi(\bar{x},\bar{\lambda}) = \alpha^{-1}\|(x-\bar{x}, 1-\bar{\lambda})\|$ and $x = \bar{x}$, we must have

$$f_\infty(\bar{x}) = f^\pi(\bar{x},0) = f^\pi(\bar{x},\bar{\lambda}) = \alpha^{-1}\|(x-\bar{x},1-\bar{\lambda})\| = 1/\alpha > 0.$$

However, from Proposition 7.2, (7.12), and $\bar{\lambda} = 0$, we obtain

$$0 \in \partial f_\infty(\bar{x}).$$

This implies that $\bar{x}$ minimizes the gauge function f_∞, meaning that $f_\infty(\bar{x}) = 0$, which is a contradiction.

The conclusion in part (i) now follows immediately from (7.12), the facts that $\bar{\lambda} > 0$ and $\bar{x} = x$, and Proposition 7.2.

We now prove part (ii). Suppose that $x \in \arg\min f$. Then $f(x) > 0$ because the optimal value of (7.1a) is positive. Let $\lambda = [1 + \alpha f(x)]^{-1} \in (0,1)$ and set $\mu = 1 - \lambda$. Then

$$f(x) = \frac{1-\lambda}{\alpha\lambda}.$$

Since $\lambda > 0$, we have $0 \in \partial f(\lambda x/\lambda)$, which together with Proposition 7.2 gives

$$(0, f(x)) = (0, -f^*(0)) \in \partial f^\pi(\lambda x, \lambda), \tag{7.13}$$

where the equality follows from a direct computation and the fact that $f(x) = \inf f$. Using this, $\lambda \in (0,1)$, and $f(x) = (1-\lambda)/(\alpha\lambda) = \mu/(\alpha(1-\mu))$, we see further that

$$(0, \alpha^{-1}\mu) \in \frac{\mu}{\alpha}\partial\|(0,1-\lambda)\| \cap (1-\mu)\partial f^\pi(\lambda x,\lambda). \tag{7.14}$$

Furthermore, we have

$$f^\pi(\lambda x, \lambda) = \lambda f(x) = \frac{1-\lambda}{\alpha} = \alpha^{-1}\|(0, 1-\lambda)\|,$$

and from (7.13) and the fact that $\lambda \in (0,1)$,

$$0 = (0, -\alpha^{-1}\mu + (1-\mu)f(x)) \in \frac{\mu}{\alpha}\partial\|(0, \lambda - 1)\| + (1-\mu)\partial f^\pi(\lambda x, \lambda),$$

showing that

$$(1-\mu)f^\pi(\lambda x, \lambda) + \frac{\mu}{\alpha}\|(0, 1-\lambda)\| = \inf_{\widehat{x},\hat{\lambda}} \max\left\{ f^\pi(\widehat{x}, \widehat{\lambda}),\ \alpha^{-1}\|(\widehat{x} - \lambda x, \widehat{\lambda} - 1)\|\right\}$$

thanks to Zălinescu [19, Corollary 2.8.15]. This, together with (7.14) and Proposition 2.2, shows that

$$(0, \alpha^{-1}\mu) \in \partial(f_1 \diamond f_2)(\lambda x, 1)$$

with f_1 and f_2 defined in (7.6). Invoking (7.7), we conclude further that $0 \in \partial \mathsf{p}_{\alpha,f}(\lambda x)$, meaning that $\lambda x \in \arg\min \mathsf{p}_{\alpha,f}$. ∎

Our next theorem states that one can obtain an optimal solution of (7.1a) by considering some "projected" fixed points of the projected polar proximal map. This is an analogue of the well-known fact that, for a proper closed convex function h, one has $\arg\min h = \{\, x \mid \mathsf{prox}_{\gamma h}(x) = x \,\}$ for any $\gamma > 0$.

THEOREM 7.4 (projected fixed points of projected polar proximal map). *Consider f as in (7.1) and let $\alpha > 0$.*

(i) *If $(x, \lambda) = \mathfrak{P}_{\alpha,\mathrm{f}}(x)$, then $\lambda > 0$ and $\lambda^{-1}x \in \arg\min f$.*

(ii) *If $x \in \arg\min f$, then there exists $\lambda > 0$ so that $(\tau x, \lambda) = \mathfrak{P}_{\alpha,\mathrm{f}}(\tau x)$, where $\tau := [1 + \alpha f(x)]^{-1}$.*

Proof. We first prove part (i). Suppose that $(x, \lambda) = \mathfrak{P}_{\alpha,\mathrm{f}}(x)$. According to Theorem 4.4(ii), there are two cases to consider.

Case 1. $f^\pi(x,\lambda) < \alpha^{-1}\|(x,\lambda) - (x,1)\|$. In particular, $\lambda \neq 1$. Using this, Zălinescu [19, Corollary 2.8.15], and the fact that (x, λ) is a minimizer, we see that

$$\begin{aligned} 0 &\in \alpha^{-1}\partial\|\cdot - (x,1)\|(x,\lambda) + \mathcal{N}((x,\lambda) \mid \operatorname{dom} f^\pi) \\ &= \frac{1}{\alpha|1-\lambda|}(0, \lambda - 1) + \mathcal{N}((x,\lambda) \mid \operatorname{dom} f^\pi). \end{aligned}$$

We use (7.3) to obtain the equivalent expression

$$\frac{1}{\alpha|1-\lambda|}(0, 1-\lambda) \in \{\, w \in [\operatorname{dom} f \times \{1\}]^- \mid \langle w, (x,\lambda)\rangle = 0 \,\}.$$

Thus, for any $y \in \operatorname{dom} f$, we have

$$0 \geq \langle (y,1), (0, 1-\lambda)\rangle \quad \text{and} \quad 0 = \langle (x,\lambda), (0, 1-\lambda)\rangle.$$

Since $\lambda \neq 1$, the first inequality above gives $\lambda > 1$, which, together with the second relation above, gives $\lambda = 0$, leading to a contradiction. Thus, this case cannot happen.

Case 2. $f^\pi(x,\lambda) = \alpha^{-1}\|(x,\lambda)-(x,1)\|$. We first claim that $\lambda \neq 1$. Suppose to the contrary that $\lambda = 1$. Then $f^\pi(x,1) = \alpha^{-1}\|(x,1)-(x,1)\| = 0$. From this and the definition of f^π, we then have $f(x) = 0$, which is impossible because the optimal value of (7.1a) is positive. Thus, $\lambda \neq 1$.

Now, using this, Zălinescu [19, Corollary 2.8.15], and the fact that (x,λ) is a minimizer, we have

$$0 = u + v$$

for some $u \in \partial(\mu^+ f^\pi)(x,\lambda)$, $v = \frac{1-\mu}{\alpha|1-\lambda|}(0,\lambda-1)$, and $\mu \in [0,1]$.

We claim that $\mu > 0$. Suppose to the contrary that $\mu = 0$. Then we have

$$0 \in \frac{1}{\alpha|1-\lambda|}(0,\lambda-1) + \mathcal{N}((x,\lambda) \mid \operatorname{dom} f^\pi).$$

A contradiction can be derived exactly as in case 1. Thus, $\mu > 0$.

Next, we claim that $\lambda > 0$. Suppose to the contrary that $\lambda = 0$. Because $\mu > 0$,

$$\mu^{-1}u \in \partial f^\pi(x,0).$$

On the other hand, we also have from $u + v = 0$ that

$$u = -v = \frac{1-\mu}{\alpha|1-\lambda|}(0,1-\lambda) = \frac{1-\mu}{\alpha}(0,1).$$

Combining the previous two displays with Proposition 7.2, we conclude that $0 \in \partial f_\infty(x)$. This implies that x minimizes the gauge function f_∞, meaning that

$$0 = f_\infty(x) = f^\pi(x,0) = \alpha^{-1}\|(x,0)-(x,1)\| = \alpha^{-1} > 0,$$

which is a contradiction. Thus, $\lambda > 0$.

Now, with $\mu > 0$, we have from $u \in \partial(\mu^+ f^\pi)(x,\lambda)$ and $u + v = 0$ that

$$\frac{u}{\mu} \in \partial f^\pi(x,\lambda) \quad \text{and} \quad u = -v = \frac{1-\mu}{\alpha|1-\lambda|}(0,1-\lambda).$$

Using this last display, $\lambda > 0$, and Proposition 7.2, we conclude that $0 \in \partial f(\lambda^{-1}x)$, as desired. This proves part (i).

We now prove part (ii). Suppose that $x \in \operatorname{arg\,min} f$. By Theorem 7.3(ii), we have $\tau x \in \operatorname{arg\,min} \mathsf{p}_{\alpha,f}$, where $\tau = [1+\alpha f(x)]^{-1}$. Now let $(\tilde{x},\lambda) = \mathfrak{P}_{\alpha,\mathrm{f}}(\tau x)$. Then Theorem 7.3(i) gives $\tilde{x} = \tau x$ and $\lambda > 0$. ∎

7.2. Projected polar proximal-point algorithm and its variants. We are motivated by the optimality conditions in Theorem 7.4 to consider a fixed-point iteration for solving (7.1a). This iteration mirrors the well-known proximal-point algorithm [12], which is a fixed-point iteration based on the (usual) proximal map. We call our algorithm the *projected polar proximal-point algorithm.*

Projected polar proximal-point algorithm ($\mathbf{P^4A}$): Fix any $\alpha > 0$, start with any x_0 and update

$$(x_{k+1},\lambda_{k+1}) = \mathfrak{P}_{\alpha,\mathrm{f}}(x_k).$$

Let $\{ (x_k, \lambda_k) \}$ be a sequence generated by P^{4}A. If $\|x_{k+1} - x_k\| \to 0$ and if (x_*, λ_*) is an accumulation point of $\{(x_k, \lambda_k)\}$, then it is routine to show that

$$(x_*, \lambda_*) = \mathfrak{P}_{\alpha,\mathrm{f}}(x_*),$$

which according to Theorem 7.4 implies that $\lambda_* > 0$ and $\lambda_*^{-1} x_* \in \arg\min f$. In the next theorem, we will show that $\|x_{k+1} - x_k\| \to 0$ under an additional assumption.

THEOREM 7.5 (convergence of P^{4}A). *Consider f as in* (7.1) *and let $\alpha > 0$. Let $\{(x_k, \lambda_k)\}$ be generated by P^{4}A. Then $\mathsf{p}_{\alpha,f}(x_{k+1}) \le \mathsf{p}_{\alpha,f}(x_k)$ for all k. If in addition there exists $\gamma > 0$ so that the function $(x, \lambda) \mapsto [f^\pi(x, \lambda)]^2 - (\gamma/2)\|x\|^2$ is convex, then $\|x_{k+1} - x_k\| \to 0$.*

Remark 7.6 (comments on the convexity condition). We give a simple sufficient condition for the convexity condition used in the hypothesis of Theorem 7.5. Let $\mathcal{C}$ be a closed convex set that does not contain the origin and let ρ be a closed gauge. For any fixed $\epsilon > 0$, consider the function $\kappa(x) := \sqrt{\rho^2(x) + \epsilon\|x\|^2}$, which is a perturbation of the gauge ρ. Then κ is a gauge function [14, Corollary 15.3.1] and κ^2 is strongly convex, say, with modulus $\gamma > 0$. If we set $f(x) := \kappa(x) + \delta_{\mathcal{C}}(x)$, then

$$f^\pi(x, \lambda) = \begin{cases} \kappa(x) + \delta_{\lambda\mathcal{C}}(x) & \text{if } \lambda > 0, \\ \kappa(x) + \delta_{\mathcal{C}_\infty}(x) & \text{if } \lambda = 0. \end{cases}$$

It is routine to show that $(x, \lambda) \mapsto [f^\pi(x, \lambda)]^2 - (\gamma/2)\|x\|^2$ is convex.

Proof. Let $(x_{k+2}, \lambda_{k+2}) = \mathfrak{P}_{\alpha,\mathrm{f}}(x_{k+1})$ and $(x_{k+1}, \lambda_{k+1}) = \mathfrak{P}_{\alpha,\mathrm{f}}(x_k)$. Then,

$$\begin{aligned} \mathsf{p}_{\alpha,f}(x_{k+1}) &\overset{\text{(a)}}{=} \max\{ f^\pi(x_{k+2}, \lambda_{k+2}),\ \alpha^{-1}\|(x_{k+2} - x_{k+1}, \lambda_{k+2} - 1)\| \} \\ &\overset{\text{(b)}}{\le} \max\left\{ f^\pi(x_{k+1}, \lambda_{k+1}),\ \alpha^{-1}\|(x_{k+1} - x_{k+1}, \lambda_{k+1} - 1)\|\right\} \\ &= \max\left\{ f^\pi(x_{k+1}, \lambda_{k+1}),\ \alpha^{-1}|\lambda_{k+1} - 1|\right\} \\ &\le \max\left\{ f^\pi(x_{k+1}, \lambda_{k+1}),\ \alpha^{-1}\|(x_{k+1} - x_k, \lambda_{k+1} - 1)\|\right\} \overset{\text{(c)}}{=} \mathsf{p}_{\alpha,f}(x_k), \end{aligned} \tag{7.15}$$

where (a) and (c) follow from the definition of $\mathsf{p}_{\alpha,f}$, and (b) follows from the definition of (x_{k+2}, λ_{k+2}) as a minimizer, so that the function value at (x_{k+2}, λ_{k+2}) is less than that at (x_{k+1}, λ_{k+1}). This proves that $\mathsf{p}_{\alpha,f}(x_{k+1}) \le \mathsf{p}_{\alpha,f}(x_k)$, as required.

Now suppose in addition that

$$(x, \lambda) \mapsto [f^\pi(x, \lambda)]^2 - (\gamma/2)\|x\|^2$$

is convex. Let $\tau = \min\{ \gamma, \alpha^{-2} \}$. Then, for each k, the function

$$[f^\pi(x, \lambda)]^2 - (\tau/2)\|x - x_k\|^2$$

is convex. Thus, the function

$$G_k(x, \lambda) := \max\left\{[f^\pi(x, \lambda)]^2, \alpha^{-2}\|(x - x_{k+1}, \lambda - 1)\|^2\right\} - (\tau/2)\|x - x_{k+2}\|^2$$

is convex. Since (x_{k+2}, λ_{k+2}) minimizes the convex function

$$(x, \lambda) \mapsto \max\left\{[f^\pi(x, \lambda)]^2, \alpha^{-2}\|(x - x_{k+1}, \lambda - 1)\|^2\right\},$$

the first-order optimality condition implies that (x_{k+2}, λ_{k+2}) also minimizes G_k. Thus,

$$\begin{aligned}
&\mathsf{p}^2_{\alpha,f}(x_{k+1}) = \max\{\,[f^\pi(x_{k+2},\lambda_{k+2})]^2, \alpha^{-2}\|(x_{k+2}-x_{k+1}, \lambda_{k+2}-1)\|^2\,\}\\
&= G_k(x_{k+2},\lambda_{k+2}) \le G_k(x_{k+1},\lambda_{k+1})\\
&= \max\left\{[f^\pi(x_{k+1},\lambda_{k+1})]^2, \alpha^{-2}\|(x_{k+1}-x_{k+1},\lambda_{k+1}-1)\|^2\right\} - (\tau/2)\|x_{k+2}-x_{k+1}\|^2\\
&\le \mathsf{p}^2_{\alpha,f}(x_k) - (\tau/2)\|x_{k+2}-x_{k+1}\|^2,
\end{aligned}$$

where the last inequality follows from (7.15). Rearranging the terms in the above inequality and summing from $k=0$ to ∞, we obtain

$$\begin{aligned}
\sum_{k=0}^{\infty}\frac{\tau}{2}\|x_{k+2}-x_{k+1}\|^2 &\le \sum_{k=0}^{\infty}(\mathsf{p}^2_{\alpha,f}(x_k) - \mathsf{p}^2_{\alpha,f}(x_{k+1}))\\
&\le \mathsf{p}^2_{\alpha,f}(x_0) < \infty,
\end{aligned}$$

which implies that $\|x_{k+1}-x_k\| \to 0$, as desired. ◻

We next describe another natural algorithm for solving (7.1a), motivated by Theorem 7.3 instead. This theorem asserts that we can recover a solution of (7.1a) by computing $\mathfrak{P}_{\alpha,\mathrm{f}}$ at a minimizer of $\mathsf{p}_{\alpha,f}$. Thus, in order to solve (7.1a), one can just minimize the projected polar envelope $\mathsf{p}_{\alpha,f}$. Specifically, we can apply a variant of the steepest-descent algorithm with line search to minimize $\mathsf{p}_{\alpha,f}$. Such an approach can be understood as a proximal-point-like algorithm in the same way that the classical proximal-point algorithm can be regarded as a steepest-descent algorithm applied to the Moreau envelope.

Projected polar envelope minimization algorithm (EMA): Start with any x_0 and $\sigma \in (0,1)$. For each k, pick $\beta_k > 0$ (so that $0 < \inf \beta_k \le \sup \beta_k < \infty$) and perform the following iteration.

1. Find the smallest integer $t \ge 0$ so that

$$\mathsf{p}_{\alpha,f}(x_k - 2^{-t}\beta_k \nabla \mathsf{p}_{\alpha,f}(x_k)) \le \mathsf{p}_{\alpha,f}(x_k) - \sigma 2^{-t}\beta_k\|\nabla \mathsf{p}_{\alpha,f}(x_k)\|^2.$$

2. Set $x_{k+1} = x_k - 2^{-t}\beta_k\nabla\mathsf{p}_{\alpha,f}(x_k)$.

We have the following convergence result concerning EMA.

THEOREM 7.7 (convergence of EMA). *Consider f as in* (7.1) *and let $\alpha > 0$. Let $\{\,x_k\,\}$ be generated by EMA. Then $\{\,x_k\,\}$ converges to a global minimizer of $\mathsf{p}_{\alpha,f}$.*

Proof. We first claim that $\mathsf{p}_{\alpha,f}(x) > 0$ for all x. Suppose to the contrary that $\mathsf{p}_{\alpha,f}(x) = 0$ for some x. Then we must have $f^\pi(\bar{x},\bar{\lambda}) = (1/\alpha)\|(\bar{x},\bar{\lambda}) - (x,1)\| = 0$, where $(\bar{x},\bar{\lambda}) \in \mathfrak{P}_{\alpha,\mathrm{f}}(x)$. Thus, $\bar{x} = x$ and $\bar{\lambda} = 1$. Consequently, we deduce that

$$0 = f^\pi(\bar{x},\bar{\lambda}) = f^\pi(x,1) = f(x),$$

contradicting the assumption that (7.1a) has a positive optimal value. Thus, $\mathsf{p}_{\alpha,f}(x) > 0$ for all x. Together with Theorem 4.4(iii) and (iv), we deduce further that $\mathsf{p}_{\alpha,f}$ is a gauge function with a continuous gradient. Moreover, using Theorem 7.3 together with the assumption that $\arg\min f \ne \emptyset$ (see (7.1)), we have $\arg\min \mathsf{p}_{\alpha,f} \ne \emptyset$. The desired conclusion now follows from these results and Iusem [11, Theorem 1]. ◻

EMA is closely related to P^{4}A. Indeed, according to Theorem 4.4(iv), $\nabla\mathsf{p}_{\alpha,f}(x)$ is a positive scaling of $x - \bar{x}$, where $(\bar{x},\bar{\lambda}) = \mathfrak{P}_{\alpha,\mathrm{f}}(x)$. Thus, one can think of EMA as a version of P^{4}A with a line-search strategy incorporated to guarantee convergence.

8. Concluding remarks. This paper continues the authors' investigation into the optimization of gauge functions and their applications [1, 9]. Gauge functions seem to appear naturally when modelling certain classes of inverse problems, including sparse optimization and machine learning. Our goal is to help establish the foundation for tools and algorithms specialized to this family of problems.

Several research avenues arise from our results and remain to be explored. The polar envelope and proximal map share many of the features of the Moreau envelope and its proximal map, and we have highlighted many of these. However, we have yet to discover an analogue for the Moreau identity, which relates the proximal map of a convex function with the map of its Fenchel conjugate. Can we similarly connect the polar proximal map of a gauge function with the map of its polar? Only the simple case is obvious: when the functions are indicators to a closed convex cone, the (Moreau) proximal and polar proximal maps coincide (cf. Example 4.2).

Along similar lines, the referees raised several important open questions that we are not yet able to answer. Does the duality between strong convexity and Lipschitz differentiability of the conjugate operation have an analogue under the polarity operation? Is it possible to specify conditions under which the regularized problem (6.4) exhibits an exactness property and recovers a solution of the unregularized problem?

The polar proximal-point algorithms that we proposed in section 7 represent a first step in developing practical algorithms, in the same way that, say, the proximal-point algorithm can be used as a springboard for many other algorithms used in practice, such as augmented Lagrangian, proximal-bundle, and proximal-gradient algorithms. What are the implementable forms of the polar proximal-point algorithms?

Acknowledgments. We extend our thanks to Alberto Seeger, who patiently answered our queries regarding his work on the max-convolution operation that laid the foundation for our investigation, and to Constantin Zălinescu for providing us with useful references. We are also grateful to the two anonymous referees for insightful comments and supplying future avenues of research.

REFERENCES

[1] A. Y. Aravkin, J. V. Burke, D. Drusvyatskiy, M. P. Friedlander, and K. MacPhee, *Foundations of gauge and perspective duality*, SIAM J. Optim., 28 (2018), pp. 2406–2434.

[2] H. Attouch, *Analyse Épigraphique*, Notes de cours de diplôme d'études approfondies, Université de Montpellier, France, 1990.

[3] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer Monogr. Math., Springer, New York, 2003.

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

[5] A. Beck and M. Teboulle, *Smoothing and first order methods: A unified framework*, SIAM J. Optim., 22 (2012), pp. 557–580.

[6] S. Becker, E. Candès, and M. Grant, *Templates for convex cone problems with applications to sparse signal recovery*, Math. Program. Comp., 3 (2011), pp. 165–218, https://doi.org/10.1007/s12532-011-0029-5.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), pp. 129–159.

[8] P. L. Combettes, *Perspective functions: Properties, constructions, and examples*, Set-Valued Var. Anal., 26 (2018), pp. 247–264.

[9] M. P. Friedlander, I. Macêdo, and T. K. Pong, *Gauge optimization and duality*, SIAM J. Optim., 24 (2014), pp. 1999–2022.

[10] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer, New York, 2001.

[11] A. N. Iusem, *On the convergence properties of the projected gradient method for convex optimization*, Comput. Appl. Math., 22 (2003), pp. 37–52.

[12] B. Martinet, *Brève communication: Régularisation d'inéquations variationnelles par approximations successives*, ESAIM Math. Model. Numer. Anal., 4 (1970), pp. 154–158.
[13] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.
[14] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[15] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, 1998.
[16] A. Seeger, *Epigrahical cones* I, J. Convex Anal., 18 (2011), pp. 1171–1196.
[17] A. Seeger and M. Volle, *On a convolution operation obtained by adding level sets: Classical and new results*, RAIRO Oper. Res., 29 (1995), pp. 131–154.
[18] C. Zălinescu, *Programare metematiča în spaţii normate infinit dimensionale*, Editura Academiei Române, Bucharest, 1998 (in Romanian).
[19] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.