

The Optimal Smoothings of Sublinear Functions and Convex Cones

Thabo Samakhoana* Benjamin Grimmer†

Abstract

This paper considers the problem of smoothing convex functions and sets, seeking the nearest smooth convex function or set to a given one. For convex cones and sublinear functions, a full characterization of the set of all optimal smoothings is given. These provide if and only if characterizations of the set of optimal smoothings for any target level of smoothness. Optimal smoothings restricting to either inner or outer approximations also follow from our theory. Finally, we apply our theory to provide insights into smoothing amenable functions given by compositions with sublinear functions and generic convex sets by expressing them as conic sections.

1 Introduction

The smoothing of nonsmooth functions is a well-established tool in nonsmooth analysis and nonsmooth optimization. Typically, smooth functions exist arbitrarily close to a given nonsmooth but continuous function. Then, a key tradeoff is balancing how smooth of a function one seeks versus how close it ought to be to the original. This smoothing task has found widespread usage:

- Moreau envelopes [1] provide a smoothing by taking an infimal convolution with a quadratic function. These have played key roles in the design and analysis of proximal and subgradient first-order methods [2, 3].
- Moreau-type smoothings have been considered in several nonconvex settings: Penot and Bougeard [4] provide infimal smoothings of functions expressed as a difference of two convex functions, for use in optimization. Lasry and Lions [5] provided smoothings for bounded uniformly continuous functions, for use in studying Hamilton-Jacobi equations. These smoothings were later generalized to a wider class of functions by [6].
- Particularly friendly smoothings arise for so-called asymptotic functions [7]. For example, one can replace minimizing a finite maximum $\max_i G_i(x)$ with the minimizing the smooth function $\eta \log(\sum_i \exp(G_i(x)/\eta))$ for some $\eta > 0$. Then, accelerated gradient methods can be applied, leading to accelerated guarantees, see [8]. Such approaches have been competitive in applications like max flows [9].
- Gaussian convolutional smoothing [10] provides a classic method of inducing smoothness by introducing noise and has found much success in zeroth-order methods, see [11] and the many references therein.
- Conditions closely related to the existence of nearby smooth functions have been used as a novel “fine-grained” measure of the complexity of nonsmooth optimization problems [12].

Smoothing techniques have also been applied to sets, but not as widely:

* Johns Hopkins University, Department of Applied Mathematics and Statistics, tsamakh1@jhu.edu

† Johns Hopkins University, Department of Applied Mathematics and Statistics, grimmer@jhu.edu

- Finding the smallest volume ellipsoid enclosing a finite set of points [13–15] provides a smoothing of their convex hull. More recently, the set interpolation theory of Luner and Grimmer [16] provides a minimal set (not necessarily an ellipsoid) of fixed smoothness that encloses the same convex hull as the solution to a second-order cone program.
- Smooth approximations of convex bodies in Banach spaces were considered in [17].
- Xu et al. [18] recently proposed an algorithm for minimization under conic constraints, indirectly smoothing the underlying cone. Their smoothing is done by smoothing the support function of a compact base of the dual cone and taking level sets of the smooth function.

In most of the existing literature, the choice of the smoothing is done in an ad hoc manner. For example, one may choose the Moreau envelope because it is smooth and it preserves minimizers. Note, however, the Moreau envelope does not generally minimize distance to the original function among all equally smooth functions. Similarly, the above-referenced max function smoothing of $z \mapsto \log(\sum_{i=1}^n \exp(z_i))$ is a 1-smooth and convex over-estimator of the convex function $z \mapsto \max\{z_1, \dots, z_n\}$, but is not the nearest over-estimator with these properties.

Our Contributions. The ad hoc nature of existing theory raises the following natural question:

How should one optimally smooth convex functions and sets?

In this paper, *we resolve this question for all sublinear functions and convex cones*, characterizing the whole Pareto frontier of smooth functions and sets optimally trading off smoothness and distance from the given function or set. Our theory at every step puts both function and set smoothing on equal footing, highlighting fundamental symmetries between these two settings.

As an application of our function smoothing theory, we obtain well-motivated smoothings for a subclass of amenable [19, Definition 10.23] functions that take the form $g = \sigma \circ G$ where σ is a sublinear function and G is a Lipschitz mapping with a Lipschitz continuous Jacobian. This class includes many nonsmooth functions that arise naturally in optimization; for example, minimizing finite maximums $\max\{G_1(x), \dots, G_n(x)\}$, maximum eigenvalue optimization $\lambda_{\max}(G(x))$, and nonlinear regression with an arbitrary norm $\|G(x)\|$.

As a concrete example, consider the nonsmooth minimization problem

$$\min_{x \in \mathbb{R}^d} \max\{G_1(x), \dots, G_n(x)\} = \min_{x \in \mathbb{R}^d} \sigma(G(x))$$

where $\sigma(z) = \max\{z_1, \dots, z_n\}$ and $G(x) = (G_1(x), \dots, G_n(x))$. Beck and Teboulle [8] proposed solving this problem by replacing $\sigma \circ G$ with the smooth approximation $f_\eta \circ G(x) = \eta \log(\sum_i \exp(G_i(x)/\eta))$ and then applying an accelerated gradient method to this smoothed problem. With an appropriate choice of η , this approach yields an ε -minimizer of $\sigma \circ G$ within $O(\sqrt{(L\varepsilon + \log(n)M^2)/\varepsilon})$ iterations, whenever each G_i is convex and G is M -Lipschitz with L -Lipschitz Jacobian. Our theory provides an optimal smoothing f_σ^{gen} (see Equation (4.17)) of $\sigma = \max$. With appropriately chosen η , minimizing our optimized smoothing $\eta f_\sigma^{\text{gen}}(G(x)/\eta)$ yields an ε -minimizer of $\sigma \circ G$ within $O(\sqrt{L\varepsilon + M^2}/\varepsilon)$ iterations. Asymptotically, this is an improvement by a factor of $1/\sqrt{\log(n)}$ ¹.

As an application of our cone smoothings, we show in Theorem 5.2 that a generic convex set can be decomposed into an intersection of an affine space and a cone that is amenable to our theory. By

¹We note that under different assumptions, the log-sum-exp approach may provide stronger guarantees since the function $\log(\sum_i \exp(z_i))$ is also 1-smooth under the l_∞ norm. This nuance is discussed further in Section 5.1.

replacing the cone with its optimal smoothings, we obtain smooth approximations to the original C . Such smoothing may be useful for constrained optimization problems where recent algorithms have been developed that benefit explicitly from the smoothness of constraint sets [16, 20, 21].

Outline. Section 2 introduces needed definitions from convex analysis, emphasizing symmetries between quantities defined on functions and sets. Section 3 presents our core definitions for a (sublinear) function and (conic) set to be smoothable. Then our main results are stated and proven in Section 4, characterizing the set of all optimal smoothings of any sublinear function or convex cone. Finally, Section 5 applies our main results to smoothing amenable functions and convex sets, illustrating potential benefits from leveraging optimal smoothings over existing ad hoc alternatives.

2 Preliminaries

We consider a finite-dimensional Euclidean space \mathcal{E} endowed with a norm $\|\cdot\|_{\mathcal{E}}$ that is induced by an inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$. We will drop the subscript on the inner product as it will always be clear from context. We identify the dual space of \mathcal{E}^* with \mathcal{E} itself in the usual way. For a linear operator $A : \mathcal{E} \rightarrow \mathcal{E}'$ between two Euclidean spaces, we denote its operator norm by $\|A\|_{\mathcal{E} \rightarrow \mathcal{E}'}$, i.e.,

$$\|A\|_{\mathcal{E} \rightarrow \mathcal{E}'} = \sup_{\|x\|_{\mathcal{E}} \leq 1} \|Ax\|_{\mathcal{E}'}.$$

We denote the convex hull of a given set C by $\text{conv } C$.

Distances and Projections. Let $C \subseteq \mathcal{E}$ be nonempty. We denote the interior and boundary of C by $\text{int } C$ and $\text{bdry } C$ respectively. We denote the Minkowski sum of two sets C and \tilde{C} by $C + \tilde{C} := \{x + \tilde{x} \mid x \in C, \tilde{x} \in \tilde{C}\}$ and the Hausdorff distance $\text{dist}(\tilde{C}, C)$ by

$$\begin{aligned} \text{dist}(x, C) &= \inf_{y \in C} \|x - y\|_{\mathcal{E}}, \\ d(\tilde{C}; C) &:= \sup_{\tilde{x} \in \tilde{C}} \text{dist}(\tilde{x}, C), \\ \text{dist}(\tilde{C}, C) &:= \max \{d(\tilde{C}; C), d(C; \tilde{C})\}. \end{aligned}$$

We denote orthogonal projection onto C by $P_C(x) = \text{argmin}\{\|x - y\|_{\mathcal{E}} \mid y \in C\}$. For any $x \in \mathcal{E}$ and $r \geq 0$, let $B(x, r) := \{y \in \mathcal{E} \mid \|x - y\|_{\mathcal{E}} \leq r\}$ denote the closed ball of radius r around x . When $x = 0$, we sometimes write B_r for $B(0, r)$. Thus $B_r + C = \{x + u \mid \|u\|_{\mathcal{E}} \leq r, x \in C\} = \{y \in \mathcal{E} \mid \text{dist}(y, C) \leq r\}$. Note that if C is closed and convex, then

$$P_{B(x, r) + C}(y) = x + P_C(y - x) + P_{B_r}(y - x - P_C(y - x)) \quad (2.1)$$

for all $x, y \in \mathcal{E}$ and $r \geq 0$. This can be easily established from the fact that $\|z\|_{\mathcal{E}}(z - P_{B_r}(z)) = (\|z\|_{\mathcal{E}} - r)_+ z$ for all $z \in \mathcal{E}$. Here, and henceforth, $(\alpha)_+ := \max\{0, \alpha\}$ for any $\alpha \in \mathbb{R}$.

Cones, Normals, and Horizons. We say a set K is a convex cone if it is closed under addition and positive rescaling. Given a generic set C , for $x \in C$, the associated normal and horizon cones of C at x are

$$\begin{aligned} N_C(x) &:= \{\zeta \in \mathcal{E} \mid \langle \zeta, y - x \rangle \leq 0 \text{ for all } y \in C\}, \\ C^\infty(x) &:= \{u \in \mathcal{E} \mid x + \alpha u \in C \text{ for all } \alpha \geq 0\}, \end{aligned}$$

respectively. C^∞ denotes the set $\bigcup_{x \in C} C^\infty(x)$. We recall that when C is convex and closed, C^∞ is convex and closed and $C^\infty = C^\infty(x)$ for all $x \in C$, see [19, Theorem 3.6].

Functions and Associated Sets. Let $f : \mathcal{E} \rightarrow (-\infty, \infty]$ be an extended valued function. We denote the Fenchel conjugate of f by f^* . The domain and epigraph are denoted by

$$\text{dom } f := \{x \in \mathcal{E} \mid f(x) < \infty\} \quad \text{and} \quad \text{epi } f := \{(x, \alpha) \in \mathcal{E} \times \mathbb{R} \mid f(x) \leq \alpha\}$$

respectively. f is proper if $\text{dom } f \neq \emptyset$. We say f is closed if $\text{epi } f$ is closed². The subdifferential of f at $x \in \text{dom } f$ is

$$\partial f(x) := \{g \in \mathcal{E} \mid (g, -1) \in N_{\text{epi } f}(x, f(x))\}.$$

The distance between two functions $f : \mathcal{E} \rightarrow \mathbb{R}$ and $g : \mathcal{E} \rightarrow \mathbb{R}$ is

$$\text{dist}(f, g) := \sup_{x \in \mathcal{E}} |f(x) - g(x)|. \quad (2.2)$$

For such functions, we write $f \geq g$ to denote that $f(x) \geq g(x)$ for all $x \in \mathcal{E}$.

Smooth Functions and Sets. For any $\beta > 0$, we say a function $f : \mathcal{E} \rightarrow \mathbb{R}$ is β -smooth if $\|\nabla f(x) - \nabla f(y)\|_{\mathcal{E}} \leq \beta \|x - y\|_{\mathcal{E}}$ for all $x, y \in \mathcal{E}$, where $\nabla f(x)$ is the gradient of f at x . The function $f(x) = \frac{\beta}{2} \|x\|_{\mathcal{E}}^2$ is the archetypal smooth function. In fact, a convex f is β -smooth if and only if

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|y - x\|_{\mathcal{E}}^2 \quad (2.3)$$

for all $x, y \in \mathcal{E}$.

For sets, we say a closed convex set C is β -smooth if $\|\zeta - \zeta'\|_{\mathcal{E}} \leq \beta \|x - x'\|_{\mathcal{E}}$ for any $x, x' \in C$ and any unit-norm normal vector $\zeta \in N_C(x)$ and $\zeta' \in N_C(x')$. The ball $B(0, 1/\beta)$ for $\beta > 0$ is the archetypal smooth set. The following analog to (2.3) holds for a closed convex set C ,

$$C \text{ is } \beta\text{-smooth} \iff B(x - \zeta/\beta, 1/\beta) \subseteq C \quad \text{for all unit norm } \zeta \in N_C(x). \quad (2.4)$$

Relating these two notions of smoothness, [20, Lemma 7] ensures

$$f \text{ is } \beta\text{-smooth} \implies \text{epi } f \text{ is } \beta\text{-smooth} \quad (2.5)$$

for all convex $f : \mathcal{E} \rightarrow \mathbb{R}$. Note that the converse to (2.5) is not true, see f_3 in Figure 1 for a counterexample. For a more detailed discussion on the smoothness of sets, see [20].

3 Smoothable Convex Functions and Sets

Our goal is to find smooth approximations of convex functions and sets. A convex function f is a β -smoothing of another convex function g if f is β -smooth and $\text{dist}(f, g) < \infty$. Similarly, a convex set S is a β -smoothing of another convex set C if S is β -smooth and $\text{dist}(S, C) < \infty$. Here, our main interest is in understanding when functions and sets admit smoothings and which smoothings minimize the distance $\text{dist}(f, g)$ as β varies. The following definition formalizes this.

Definition 3.1. A convex function $g : \mathcal{E} \rightarrow \mathbb{R}$ is (λ, Δ) -smoothable if for any $\beta > \Delta$, there exists a β -smooth convex function $f : \mathcal{E} \rightarrow \mathbb{R}$ such that $\text{dist}(f, g) \leq \frac{\lambda}{\beta - \Delta}$. Any such f is called a β -smoothing of g with parameters (λ, Δ) or a (λ, Δ, β) -smoothing for short.

²We always take $\mathcal{E} \times \mathbb{R}$ as a Euclidean space with inner product $\langle (x, \alpha), (x', \alpha') \rangle = \langle x, x' \rangle + \alpha\alpha'$.

Beck and Teboulle [8, Definition 2.1] introduced a very similar notion of function smoothability, motivating our definition. The above definition differs from [8, Definition 2.1] in two ways: (i) we require the smooth approximation g to be globally smooth and (2) we only require bounds on $\text{dist}(f, g)$ rather than enforcing separate tolerances bounding the difference above $g - f$ and the difference below $f - g$. We introduce the following analogous definition for sets.

Definition 3.2. A convex set C is (λ, Δ) -smoothable if for any $\beta > \Delta$, there exists a β -smooth convex set S such that $\text{dist}(S, C) \leq \frac{\lambda}{\beta - \Delta}$. Any such S is called a β -smoothing of C with parameters (λ, Δ) , or a (λ, Δ, β) -smoothing for short.

Note that not all functions are smoothable. Consider, for instance, $g(x) = x^4$. Since every β -smooth function f is upper bounded by $f(0) + f'(0)x + \frac{\beta}{2}x^2$, the difference $g - f$ must diverge as x grows, making $\text{dist}(f, g) = \infty$. Similar reasoning also establishes the necessity of restricting to $\beta > \Delta$ in the above definition. Consider, for instance $g(x) = \frac{\Delta}{2}x^2 + |x|$ for any $\Delta > 0$. The above reasoning establishes a β -smoothing f with $\text{dist}(f, g) < \infty$ can only exist for $\beta > \Delta$. In contrast, every convex set is smoothable. Indeed, for any closed, convex C and $\beta > 0$, the set $S := B_{1/\beta} + C$ is β -smooth and satisfies $\text{dist}(C, S) \leq 1/\beta$. The fact that $B_{1/\beta} + C$ is smooth is a consequence of a more general result: The (Minkowski) sum of two closed convex sets is smooth whenever one of them is smooth, see [20, Lemma 9].

3.1 Smoothings of Sublinear Functions and Convex Cones

By restricting to sublinear functions σ and convex cones K , the above notions of smoothability can be simplified, setting $\Delta = 0$ without loss of generality. To see this, consider the following simple rescaling operations: for any function $f : \mathcal{E} \rightarrow \mathbb{R}$, set C , and $\eta > 0$, define

$$[f]_\eta(x) = \eta f(x/\eta) \quad [C]_\eta = \eta C = \{\eta x \mid x \in C\}. \quad (3.1)$$

Note that $\text{epi } [f]_\eta = \eta \cdot \text{epi } f = [\text{epi } f]_\eta$. Therefore, the functional and set rescalings above are one and the same when viewing functions equivalently through their epigraphs.

Leveraging positive homogeneity, the following lemma characterizes the smoothness of $[f]_\eta$ and $[C]_\eta$ and how their distance to a given sublinear σ and cone K scales with η .

Lemma 3.3. For any $\eta > 0$, function f and sublinear σ , and set C and cone K ,

$$\text{dist}([f]_\eta, \sigma) = \eta \cdot \text{dist}(f, \sigma) \quad \text{and} \quad \text{dist}([C]_\eta, K) = \eta \cdot \text{dist}(C, K). \quad (3.2)$$

In addition, if f and C are β -smooth, then $[f]_\eta$ and $[C]_\eta$ are β/η -smooth respectively.

Proof. We have

$$\text{dist}([f]_\eta, \sigma) = \sup_{x \in \mathcal{E}} |\eta f(x/\eta) - \sigma(x)| = \sup_{x \in \mathcal{E}} \eta |f(x/\eta) - \sigma(x/\eta)| = \eta \text{dist}(f, \sigma)$$

where the third equality uses that $\lambda\sigma(x) = \sigma(\lambda x)$ whenever $\lambda \geq 0$. Similarly

$$\text{dist}([C]_\eta, K) = \text{dist}(\eta C, K) = \text{dist}(\eta C, \eta K) = \eta \text{dist}(C, K)$$

where the third equality follows because $\lambda K = K$ whenever $\lambda > 0$. For smoothness, we have

$$\|\nabla [f]_\eta(x) - \nabla [f]_\eta(y)\|_{\mathcal{E}} = \|\nabla f(x/\eta) - \nabla f(y/\eta)\|_{\mathcal{E}} \leq (\beta/\eta) \|x - y\|_{\mathcal{E}}$$

where the equality follows by the chain rule and the inequality by β -smoothness of f . Similarly, for any unit norm $\zeta_1 \in N_{[C]_\eta}(x)$ and $\zeta_2 \in N_{[C]_\eta}(y)$ we have

$$\|\zeta_1 - \zeta_2\|_{\mathcal{E}} \leq \beta \|x/\eta - y/\eta\|_{\mathcal{E}} = (\beta/\eta) \|x - y\|_{\mathcal{E}}$$

where the inequality follows from β -smoothness of C and the identity $N_{[C]_\eta}(z) = N_C(z/\eta)$. \square

From Lemma 3.3, it is clear that if a sublinear function σ or a cone K is (λ, Δ) -smoothable, then it must be $(\lambda, 0)$ -smoothable. Hence, going forward in our study of optimal smoothings of sublinear functions and convex cones, it suffices to fix $\Delta = 0$. In this case, we will abuse notation and say σ or K is λ -smoothable to denote $(\lambda, 0)$ -smoothability.

3.2 Inner and Outer Smoothings

In Definitions 3.1 and 3.2, there is no distinction on whether the approximation is from above or below, or inside or outside. It can sometimes be useful in practice to make such a distinction. This leads to the following definitions for function and set smoothing from inside and from outside. When applying notions of “inner”/“outer” approximations to functions, we take the convention of identifying each function with its epigraph. Hence, an inner approximation f of g is a function with $\text{epi } f \subseteq \text{epi } g$. That is, $f \geq g$.

Definition 3.4. A convex function g is inner (or outer) (λ, Δ) -smoothable if for any $\beta > \Delta$, there exists a (λ, Δ, β) -smoothing f such that $f \geq g$ (or $f \leq g$).

Definition 3.5. A convex set C is inner (or outer) (λ, Δ) -smoothable if for any $\beta > \Delta$, there exists a (λ, Δ, β) -smoothing S such that $S \subseteq C$ (or $S \supseteq C$).

As noted in the previous subsection, for sublinear functions and convex cones, Δ can be set to zero in all the above definitions without loss of generality. In such settings, we abuse notation and omit $\Delta = 0$ from our statements.

3.3 Example: Smoothing the Euclidean Norm and the Second Order Cone

To illustrate these definitions, we briefly consider smoothing the standard Euclidean norm $\sigma(x) = \|x\|_2$ and the second-order cone $K := \text{epi } \|\cdot\|_2$. First, we focus on smoothing σ . Consider the following five candidate smoothings of the Euclidean norm

$$\begin{aligned} f_1(x) &:= \begin{cases} \frac{1}{2} \|x\|_2^2 & \text{if } \|x\|_2 \leq 1 \\ \|x\|_2 - \frac{1}{2} & \text{otherwise} \end{cases} & f_2(x) &:= f_1(x) + \frac{1}{4} \\ f_3(x) &:= \begin{cases} 2\sqrt{2} - \sqrt{8 - \|x\|_2^2} & \text{if } \|x\|_2 \leq 2 \\ \|x\|_2 - 2(2 - \sqrt{2}) & \text{otherwise} \end{cases} & f_4(x) &:= f_3(x) + 6\sqrt{2} - 8 \\ f_5(x) &:= \sqrt{1 + \|x\|_2^2} - 1 \end{aligned}$$

where f_1 and f_2 are the Moreau envelope of σ (known as a Huber function) and a translation of it, f_3 and f_4 are functions whose epigraph is a translation of $\text{epi } \sigma + B_{2\sqrt{2}}$, and f_5 is an ad hoc choice avoiding piecewise definitions. One can verify that f_1, f_3, f_5 are all outer smoothings while f_2 and f_4 are neither inner nor outer smoothings.

For each $i = 1, 2, \dots, 5$, Figure 1 computes constants for the β_i -smoothness of each f_i and the distance $D_i = \text{dist}(f_i, \sigma)$. Since $\sigma = \|\cdot\|_2$ is sublinear, Lemma 3.3 ensures $[f_i]_\eta$ is β_i/η smoothing

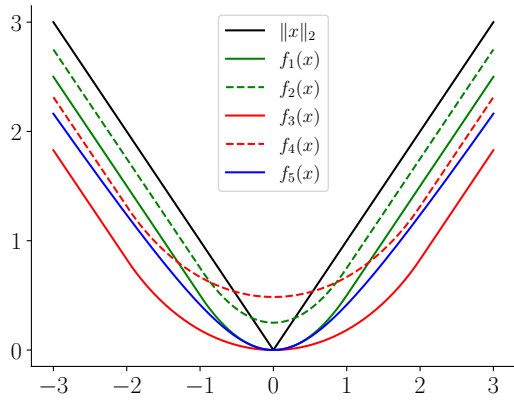


Figure 1: Left: Five candidate smoothings of $\|x\|_2$ in 1D. Right: Smoothness and distance bounds, and the implied smoothability bound for the two-norm and second-order cone from each candidate.

with distance from σ of ηD_i . As a result, each f_i proves σ is $\lambda_i = \beta_i D_i$ -smoothable. Among these choices, f_2 provides the best smoothability constant in general and f_1 provides the best constant among the outer smoothings. Our Theorems 4.1 and 4.5 establish that, in fact, these are the unique optimal smoothing and outer smoothing of the Euclidean norm among all 1-smooth functions.

Each of these function approximations provides a smooth set approximating the second-order cone $K = \text{epi } \sigma$ given by $S_i = \text{epi } f_i$. Each of these sets has some associated smoothness constant β'_i and distance $D'_i = \text{dist}(S_i, K)$, also computed in Figure 1. Again by Lemma 3.3, we know $[S_i]_\eta$ is β'_i/η smoothing with distance from K of ηD_i . Hence, each S_i proves the second-order cone K is smoothable with constant $\lambda'_i = \beta'_i D'_i$. Among these, S_4 provides the best smoothability constant in general and S_3 provides the best constant among the outer smoothings. Our Theorems 4.2 and 4.6 establish that these are the unique optimal smoothings.

After developing our theory for generic sublinear functions and convex cones, more interesting examples are considered in Section 4.2. Unlike our simple example here, optimal smoothings in general are not unique; in such settings, we provide a characterization of all optimal smoothings. The fact that the Moreau envelope of the two-norm plus a constant was an optimal function smoothing is shared among norms in general, but does not always provide optimal smoothings for sublinear functions.

4 Optimal Smoothings of Sublinear Functions and Cones

As our main result, we characterize the set of all optimal smoothings of any sublinear function or convex cone. Recall that $\sigma : \mathcal{E} \rightarrow \mathbb{R}$ is sublinear if

$$\sigma(\lambda x) = \lambda \sigma(x) \text{ if } \lambda \geq 0, \quad \sigma(x + y) \leq \sigma(x) + \sigma(y)$$

and $K \subseteq \mathcal{E}$ is a convex cone if

$$\lambda x \in K \text{ if } x \in K, \lambda \geq 0, \quad x + y \in K \text{ if } x, y \in K.$$

For the rest of the paper, σ will denote a sublinear function and K a closed, convex cone, with a nonempty interior such that $K \neq \mathcal{E}$. We define optimal smoothings of functions and sets as follows.

Definition 4.1. A convex function $f : \mathcal{E} \rightarrow \mathbb{R}$ is an optimal β -smoothing of $g : \mathcal{E} \rightarrow \mathbb{R}$ if f is β -smooth and $\text{dist}(g, f) \leq \text{dist}(g, \tilde{f})$ for all β -smooth, convex $\tilde{f} : \mathcal{E} \rightarrow \mathbb{R}$.

Definition 4.2. A closed convex set S is an optimal β -smoothing of another set C if S is β -smooth and $\text{dist}(C, S) \leq \text{dist}(C, \tilde{S})$ for all β -smooth, convex \tilde{S} .

Recall that as an immediate consequence of Lemma 3.3, it suffices to characterize all optimal smoothings of σ and K at the fixed smoothness level $\beta = 1$. To construct such smoothings, we use the following key objects, which we call the (“functional” for σ or “conic” for K) core, center, and width, respectively:

$$C_\sigma := \left\{ (x, r) \in \mathcal{E} \times \mathbb{R} \mid (x, r) + \text{epi } \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \subseteq \text{epi } \sigma \right\}, \quad C_K := \{x \in \mathcal{E} \mid x + B(0, 1) \subseteq K\},$$

$$(x_\sigma, r_\sigma) := \underset{(x, r) \in C_\sigma}{\text{argmin}} \, r + \frac{1}{2} \|x\|_{\mathcal{E}}^2,$$

$$x_K := \underset{x \in C_K}{\text{argmin}} \, \|x\|_{\mathcal{E}},$$

$$w_\sigma := r_\sigma + \frac{1}{2} \|x_\sigma\|_{\mathcal{E}}^2,$$

$$w_K := \|x_K\|_{\mathcal{E}} - 1.$$

We claim the functional and conic cores are both closed and convex. The following pair of lemmas establishes this by giving equivalent dual formulas for both objects as a maximum of affine functions and as an intersection of halfspaces. In particular, the functional core C_σ is the epigraph of the following closed, convex function:

$$\varrho_\sigma(x) := \max_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_{\mathcal{E}}^2. \quad (4.1)$$

Lemma 4.3. For any sublinear function σ , $C_\sigma = \text{epi } \varrho_\sigma$.

Proof. Observe that it suffices to show that ϱ_σ equals $\sup_{y \in \mathcal{E}} \sigma(y) - \frac{1}{2} \|x - y\|_{\mathcal{E}}^2$. This is sufficient since for any $(x, r) \in \mathcal{E} \times \mathbb{R}$,

$$\begin{aligned} (x, r) \in C_\sigma &\iff r + \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 \geq \sigma(y) \text{ for all } y \in \mathcal{E} \iff r \geq \sup_{y \in \mathcal{E}} \sigma(y) - \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 \\ &\iff (x, r) \in \text{epi } \varrho_\sigma \end{aligned}$$

where the first equivalence follows by definition of C_σ while the last one uses the claimed equality.

Note that the supremum $\sup_{y \in \mathcal{E}} \sigma(y) - \frac{1}{2} \|x - y\|_{\mathcal{E}}^2$ will always be attained since the function has compact level sets. Then any \hat{y} attaining this must have $0 \in \hat{y} - x + \partial_P(-\sigma)(\hat{y})$ by [22, Corollary 1.2.7] and [22, Proposition 1.2.11], where ∂_P denotes the proximal subdifferential. Since σ is convex, this implies $\hat{y} = x + \nabla\sigma(\hat{y})$ see [22, Exercise 1.2.8 (c)]. Therefore $\sup_{y \in \mathcal{E}} \sigma(y) - \frac{1}{2} \|x - y\|_{\mathcal{E}}^2$ is equal to

$$\max_{\zeta \in \partial\sigma(x+\zeta)} \sigma(x+\zeta) - \frac{1}{2} \|x - (x+\zeta)\|_{\mathcal{E}}^2 = \max_{\zeta \in \partial\sigma(x+\zeta)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_{\mathcal{E}}^2 = \max_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_{\mathcal{E}}^2 = \varrho_\sigma(x)$$

where the first equality follows because $\sigma(y) = \langle \zeta, y \rangle$ whenever $\zeta \in \partial\sigma(y)$ by the Fenchel-Young equality, and the second equality follows because any maximizer $\hat{\zeta}$ of the problem on the right satisfies $\hat{\zeta} \in \partial\sigma(x + \hat{\zeta})$ by the necessary condition $x + \hat{\zeta} \in N_{\partial\sigma(0)}(\hat{\zeta})$, see [22, Exercise 1.2.12]. \square

Lemma 4.4. For any convex cone K , $C_K = \{x \in \mathcal{E} \mid \langle \zeta, x \rangle \leq -1, \zeta \in N_K(0), \|\zeta\|_{\mathcal{E}} = 1\}$.

Proof. Let $x \in C_K$. Then for any $\zeta \in N_K(0)$ with $\|\zeta\|_{\mathcal{E}} = 1$, we have

$$\langle \zeta, x \rangle = -1 + \langle \zeta, x + \zeta \rangle \leq -1$$

where the first equality follows because $\langle \zeta, \zeta \rangle = 1$ and the inequality follows because $x \in C_K$ implies $x + \zeta \in K$ since $\|\zeta\|_{\mathcal{E}} \leq 1$. Now suppose $\sup_{\zeta \in N_K(0), \|\zeta\|_{\mathcal{E}}=1} \langle \zeta, x \rangle \leq -1$. Then, for any u with $\|u\|_{\mathcal{E}} \leq 1$, we have

$$\sup_{\zeta \in N_K(0), \|\zeta\|_{\mathcal{E}}=1} \langle \zeta, x + u \rangle \leq \|u\|_{\mathcal{E}} + \sup_{\zeta \in N_K(0), \|\zeta\|_{\mathcal{E}}=1} \langle \zeta, x \rangle \leq 0$$

where the first inequality follows from the Cauchy-Schwarz inequality and the last follows by assumption. Since K is a closed convex cone, this implies $x + u \in K$. Since u was arbitrary, it follows that $x \in C_K$. \square

From Lemma 4.4, it is immediate that the conic center x_K , defined above as the projection onto C_K , exists and is unique. Similarly, from Lemma 4.3, it follows that $\min_{(x,r) \in C_\sigma} r + \frac{1}{2} \|x\|_{\mathcal{E}}^2$ has a unique solution given by

$$x_\sigma = \operatorname{argmin}_{x \in \mathcal{E}} \varrho_\sigma(x) + \frac{1}{2} \|x\|_{\mathcal{E}}^2 \quad \text{and} \quad r_\sigma = \varrho_\sigma(x_\sigma) \quad (4.2)$$

since

$$(x_\sigma, r_\sigma) = \operatorname{argmin}_{(x,r) \in \operatorname{epi} \varrho_\sigma} r + \frac{1}{2} \|x\|_{\mathcal{E}}^2 = \operatorname{argmin}_{(x,r), r = \varrho_\sigma(x)} \varrho_\sigma(x) + \frac{1}{2} \|x\|_{\mathcal{E}}^2.$$

Noting $\varrho_\sigma(x) + \frac{1}{2} \|x\|_{\mathcal{E}}^2$ is 1-strongly convex, this minimization has a unique minimizer.

Lastly, we note that our conic width w_K is closely related to the conic width of Xiong and Freund [23, Section 1.3], which is an important quantity in the analysis of conic programming algorithms (see [23, 24]). In fact, it can easily be shown that their conic width is equal to $1/(w_K + 1) = 1/\|x_K\|_{\mathcal{E}}$.

4.1 Characterization of All Optimal Smoothings

We find that the set of optimal smoothings of a given sublinear σ always takes the structured form of all β -smooth convex functions lying pointwise between two extremal smoothings. In particular, we consider the following minimal and maximal 1-smoothings:

$$f_\sigma^{\text{gen}} = \left[(r_\sigma + \sigma(\cdot - x_\sigma)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma/2, \quad F_\sigma^{\text{gen}} = \left[\varrho_\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma/2 \quad (4.3)$$

where $f \square g$ denotes the infimal convolution between two functions $f : \mathcal{E} \rightarrow \mathbb{R}$ and $g : \mathcal{E} \rightarrow \mathbb{R}$, i.e.

$$(f \square g)(x) = \inf_{y \in \mathcal{E}} f(y) + g(x - y).$$

Similarly, for a closed convex cone K , to characterize all smoothings, we consider the following minimal and maximal 1-smoothings:

$$s_K^{\text{gen}} = \frac{1}{1 + \frac{w_K}{2}} \left[(x_K + K) + B(0, 1 + \frac{w_K}{2}) \right], \quad S_K^{\text{gen}} = \frac{1}{1 + \frac{w_K}{2}} \left[C_K + B(0, 1 + \frac{w_K}{2}) \right]. \quad (4.4)$$

The following theorems establish for functions and sets, respectively, that the smoothings above provide the smallest and largest optimal smoothings of σ and K and suffice to characterize the set of all optimal smoothings. We defer proofs of these main results to Section 4.3.

Theorem 4.1. A sublinear function σ is λ -smoothable if and only if $\lambda \geq w_\sigma/2$. In particular, for any target smoothness level $\beta > 0$, a β -smooth convex f is an optimal β -smoothing of σ if and only if $[F_\sigma^{\text{gen}}]_{1/\beta} \leq f \leq [f_\sigma^{\text{gen}}]_{1/\beta}$.

Theorem 4.2. A convex cone $K \subset \mathcal{E}$ with a nonempty interior is λ -smoothable if and only if $\lambda \geq \frac{w_K}{2+w_K}$. In particular, for any target smoothness level $\beta > 0$, a β -smooth convex S is an optimal β -smoothing of K if and only if $[s_K^{\text{gen}}]_{1/\beta} \subseteq S \subseteq [S_K^{\text{gen}}]_{1/\beta}$.

4.1.1 Characterizations for Optimal Inner and Outer Smoothings

Definition 4.5. A convex function $f : \mathcal{E} \rightarrow \mathbb{R}$ is an optimal inner β -smoothing of $g : \mathcal{E} \rightarrow \mathbb{R}$ if f is β -smooth, $f \geq g$, and $\text{dist}(g, f) \leq \text{dist}(g, \tilde{f})$ for all β -smooth, convex $\tilde{f} : \mathcal{E} \rightarrow \mathbb{R}$ satisfying $\tilde{f} \geq g$. Analogously, a β -smooth convex $f \leq g$ is an optimal outer β -smoothing of g if $\text{dist}(g, f) \leq \text{dist}(g, \tilde{f})$ for all β -smooth, convex $\tilde{f} \leq g$.

Definition 4.6. A closed convex set S is an optimal inner β -smoothing of another set C if S is β -smooth, $S \subseteq C$, and $\text{dist}(C, S) \leq \text{dist}(C, \tilde{S})$ for all β -smooth, convex $\tilde{S} \subseteq \mathcal{E}$ satisfying $\tilde{S} \subseteq C$. Analogously, a β -smooth, closed convex $S \supseteq C$ is an optimal outer β -smoothing of C if $\text{dist}(C, S) \leq \text{dist}(C, \tilde{S})$ for all β -smooth, closed convex $\tilde{S} \supseteq C$.

For any sublinear σ , we find that all optimal inner 1-smoothings of σ lie between the following functions, defined as infimal convolutions

$$f_\sigma^{\text{in}} = (r_\sigma + \sigma(\cdot - x_\sigma)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2, \quad F_\sigma^{\text{in}} = \varrho_\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2. \quad (4.5)$$

For any convex cone K , we find that all optimal inner 1-smoothings lie between the following sets, defined as Minkowski sums

$$s_K^{\text{in}} = (x_K + K) + B(0, 1), \quad S_K^{\text{in}} = C_K + B(0, 1). \quad (4.6)$$

Theorems 4.3 and 4.4 formalize this and are proven in Section 4.3.

Theorem 4.3. Every sublinear function σ is inner λ -smoothable if and only if $\lambda \geq w_\sigma$. In particular, for any target smoothness level $\beta > 0$, a β -smooth convex f is an optimal inner β -smoothing of σ if and only if $[F_\sigma^{\text{in}}]_{1/\beta} \leq f \leq [f_\sigma^{\text{in}}]_{1/\beta}$.

Theorem 4.4. Every convex cone $K \subset \mathcal{E}$ with a nonempty interior is inner λ -smoothable if and only if $\lambda \geq w_K$. In particular, for any target smoothness level $\beta > 0$, a β -smooth convex S is an optimal inner β -smoothing of K if and only if $[s_K^{\text{in}}]_{1/\beta} \subseteq S \subseteq [S_K^{\text{in}}]_{1/\beta}$.

Similarly, the set of optimal outer 1-smoothings of σ lies between the following functions

$$f_\sigma^{\text{out}} = \left[(r_\sigma + \sigma(\cdot - x_\sigma)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma, \quad F_\sigma^{\text{out}} = \left[\varrho_\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma, \quad (4.7)$$

and the set of optimal outer 1-smoothings of K lies between the following sets

$$s_K^{\text{out}} = \frac{1}{1+w_K} [(x_K + K) + B(0, 1 + w_K)], \quad S_K^{\text{out}} = \frac{1}{1+w_K} [C_K + B(0, 1 + w_K)]. \quad (4.8)$$

Theorems 4.5 and 4.6 formalize this and are proven in Section 4.3.

Theorem 4.5. Every sublinear function σ is outer λ -smoothable if and only if $\lambda \geq w_\sigma$. In particular, for any target smoothness level $\beta > 0$, a β -smooth convex f is an optimal outer β -smoothing if and only if $[F_\sigma^{\text{out}}]_{1/\beta} \leq f \leq [f_\sigma^{\text{out}}]_{1/\beta}$.

Theorem 4.6. Every convex cone $K \subset \mathcal{E}$ with a nonempty interior is outer λ -smoothable if and only if $\lambda \geq \frac{w_K}{1+w_K}$. In particular, for any target smoothness level $\beta > 0$, a β -smooth convex S is an optimal outer β -smoothing of K if and only if $[s_K^{\text{out}}]_{1/\beta} \subseteq S \subseteq [S_K^{\text{out}}]_{1/\beta}$.

4.1.2 Two Remarks on Optimal Smoothings First, we explicitly highlight the parallels between the optimal function and set smoothings. Recall that (see [19, Exercice 1.28]) strong convexity of $f + \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2$ guarantees that

$$\text{epi } f \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 = \text{epi } f + \text{epi } \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \quad \text{for all convex } f : \mathcal{E} \rightarrow \mathbb{R}. \quad (4.9)$$

Noting that $\text{epi } \sigma$ is a convex cone, (4.9) reveals the following symmetry between the optimal inner smoothings of sublinear functions and convex cones. Observe that

$$\begin{aligned} \text{epi } f_{\sigma}^{\text{in}} &= ((x_{\sigma}, r_{\sigma}) + \text{epi } \sigma) + \text{epi } \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2, & \text{epi } F_{\sigma}^{\text{in}} &= C_{\sigma} + \text{epi } \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2, \\ s_K^{\text{in}} &= (x_K + K) + B(0, 1), & S_K^{\text{in}} &= C_K + B(0, 1). \end{aligned}$$

The other optimal sets are generated from these inner smoothings. For functions, a simple translation gives the other optimal smoothings, as

$$\begin{aligned} f_{\sigma}^{\text{gen}} &= f_{\sigma}^{\text{in}} - w_{\sigma}/2, & F_{\sigma}^{\text{gen}} &= F_{\sigma}^{\text{in}} - w_{\sigma}/2, \\ f_{\sigma}^{\text{out}} &= f_{\sigma}^{\text{in}} - w_{\sigma}, & F_{\sigma}^{\text{out}} &= F_{\sigma}^{\text{in}} - w_{\sigma}. \end{aligned} \quad (4.10)$$

The parallel transformation for sets is a Minkowski addition together with a rescaling, giving

$$\begin{aligned} s_K^{\text{gen}} &= \frac{1}{1 + \frac{w_K}{2}} [s_K^{\text{in}} + B(0, \frac{w_K}{2})], & S_K^{\text{gen}} &= \frac{1}{1 + \frac{w_K}{2}} [S_K^{\text{in}} + B(0, \frac{w_K}{2})], \\ s_K^{\text{out}} &= \frac{1}{1 + w_K} [s_K^{\text{in}} + B(0, w_K)], & S_K^{\text{out}} &= \frac{1}{1 + w_K} [S_K^{\text{in}} + B(0, w_K)]. \end{aligned} \quad (4.11)$$

The reformulations in (4.10) and (4.11) are easy to verify from definitions.

Second, we note that our theory provides an exact characterization of when a unique optimal smoothing exists. For both the functional and conic setting, all the smoothings are unique if and only if the core is the translation of the original with the origin placed at the center:

$$\sigma \text{ has unique smoothings} \iff \varrho_{\sigma} = r_{\sigma} + \sigma(\cdot - x_{\sigma}), \quad (4.12)$$

$$K \text{ has unique smoothings} \iff C_K = x_K + K. \quad (4.13)$$

4.2 Some Families of Optimal Smoothings

Before proving our main theorems in the following subsection, here we apply them to several important, well-studied nonsmooth functions and sets. First, applying our theory to sublinear functions, we present the optimal smoothings of the ReLU function, an optimal smoothing of any norm based on its Moreau envelope (not necessarily unique), and optimal, computationally tractable smoothings of the maximum and maximum eigenvalue functions. Second, applying our theory to convex cones, we present the unique optimal smoothings of the nonnegative, second-order, and semidefinite cone and numerically identify the set of all optimal smoothings of the exponential cone.

For these examples, the following lemma is useful, giving a formula for the infimal convolution with $\frac{1}{2} \|\cdot\|_{\mathcal{E}}^2$ (i.e., the Moreau envelope) of any sublinear function. The minimal optimal smoothings $f_{\sigma}^{\text{gen}}, f_{\sigma}^{\text{in}}, f_{\sigma}^{\text{out}}$ are all translations of such functions. (Note the maximal optimal smoothings $F_{\sigma}^{\text{gen}}, F_{\sigma}^{\text{in}}, F_{\sigma}^{\text{out}}$ generally are not translations of Moreau envelopes of sublinear functions.)

Lemma 4.7. *For any sublinear function σ and $x \in \mathcal{E}$,*

$$(\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2)(x) = \frac{1}{2} \|x\|_{\mathcal{E}}^2 - \frac{1}{2} \text{dist}(x, \partial\sigma(0))^2. \quad (4.14)$$

Proof. First, we assert that

$$\operatorname{argmin}_{y \in \mathcal{E}} \sigma(y) + \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 = x - P_{\partial\sigma(0)}(x). \quad (4.15)$$

Let $x^+ = x - P_{\partial\sigma(0)}(x)$. It suffices to show that $x - x^+ \in \partial\sigma(x^+)$ as this is the necessary and sufficient condition to attain the minimum in (4.15). For any $\zeta \in \partial\sigma(0)$, we have

$$\langle x^+, \zeta \rangle = \langle x - P_{\partial\sigma(0)}(x), \zeta \rangle \leq \langle x - P_{\partial\sigma(0)}(x), P_{\partial\sigma(0)}(x) \rangle = \langle x^+, x - x^+ \rangle$$

where the inequality follows by normality conditions. Since σ is sublinear, this implies $x - x^+ \in \partial\sigma(x^+)$ and hence (4.15). Then the claimed general formula for the Moreau envelope of σ follows as

$$\begin{aligned} (\sigma \square \tfrac{1}{2} \|\cdot\|_{\mathcal{E}}^2)(x) &= \inf_{y \in \mathcal{E}} \sigma(y) + \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 \\ &= \sigma(x - P_{\partial\sigma(0)}(x)) + \frac{1}{2} \|P_{\partial\sigma(0)}(x)\|_{\mathcal{E}}^2 \\ &= \langle x - x_{\sigma} - P_{\partial\sigma(0)}(x), P_{\partial\sigma(0)}(x) \rangle + \frac{1}{2} \|P_{\partial\sigma(0)}(x)\|_{\mathcal{E}}^2 \\ &= \frac{1}{2} \|x\|_{\mathcal{E}}^2 - \frac{1}{2} \|x - P_{\partial\sigma(0)}(x)\|_{\mathcal{E}}^2 \end{aligned}$$

where the second equality follows by (4.15), the third because $P_{\partial\sigma(0)}(x) \in \partial\sigma(x - P_{\partial\sigma(0)}(x))$ by normality conditions, and the last is just completing the square. Since $\operatorname{dist}(x, \partial\sigma(0)) = \|x - x_{\sigma} - P_{\partial\sigma(0)}(x)\|_{\mathcal{E}}$, the claimed formula follows. \square

4.2.1 The Optimal Smoothings of the ReLU Function As a simple first example of calculating optimal smoothing with our theory, consider the ReLU (rectified linear unit) function denoted $\sigma(x) := \max\{0, x\}$ defined with $x \in \mathcal{E} = \mathbb{R}$. A wide range of ad hoc smoothings of this function have been used in the machine learning literature [25–28]. Our theory provides principled alternatives. Noting that $\partial\sigma(0) = \operatorname{conv}\{0, 1\}$, the functional core can be computed as

$$\varrho_{\sigma}(x) = \max_{\zeta \in \partial\sigma(0)} \zeta x + \frac{1}{2} \zeta^2 = \max_{\zeta \in \{0, 1\}} \zeta x + \frac{1}{2} \zeta^2 = \max \left\{ 0, x + \frac{1}{2} \right\} = \sigma(x + 1/2).$$

The functional center is given by $(x_{\sigma}, r_{\sigma}) = (-1/2, 0)$ and the functional width is $w_{\sigma} = 1/8$. Observing that $\varrho_{\sigma} = r_{\sigma} + \sigma(\cdot - x_{\sigma})$, (4.12) ensures that the ReLU function has a unique optimal smoothing. In particular, from Theorem 4.1, calculating an infimal convolution of $\frac{1}{2} \|\cdot\|_{\mathcal{E}}^2$ with ϱ_{σ} gives the unique optimal 1-smoothing as

$$f_{\sigma}^{\text{gen}}(x) = \begin{cases} -\frac{1}{16} & \text{if } x < -1/2 \\ \frac{1}{2}(x + \frac{1}{2})^2 - \frac{1}{16} & \text{if } -1/2 \leq x \leq 1/2 \\ x - \frac{1}{16} & \text{otherwise.} \end{cases} \quad (4.16)$$

The error of approximation is $\operatorname{dist}(f_{\sigma}^{\text{gen}}, \sigma) = \frac{1}{2} w_{\sigma} = 1/16$. More generally, $\frac{1}{\beta} f_{\sigma}^{\text{gen}}(\beta x)$ gives the optimal β -smoothing for any $\beta > 0$, with approximation error of $\frac{1}{16\beta}$.

4.2.2 An Optimal Smoothing for any Norm For any norm $\sigma = \|\cdot\|$ (not necessarily the Euclidean norm), we find that the function center is very structured. Namely, $x_\sigma = 0$ and $w_\sigma = \varrho_\sigma(0)$. From this, one can conclude that the Moreau envelope of the norm, $\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2$, provides an optimal outer β -smoothing. From this optimal outer smoothing, optimal smoothings in general and inner smoothings are given by appropriately adding constants to this envelope via (4.10).

Proposition 4.8. *For any norm σ and $\beta > 0$, the Moreau Envelope $\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2$ is an optimal outer β -smoothing of σ . An optimal general smoothing and inner smoothing is given by $(\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2) + \beta w_\sigma/2$ and $(\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2) + \beta w_\sigma$, respectively, where $w_\sigma = \max_{\zeta \in \partial\sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2$.*

Proof. We will show that $f_\sigma^{\text{out}} = \sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2$. Noting that $\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2 = \left[\sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \right]_{1/\beta}$ by positive homogeneity of σ , this will prove optimality of $\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2$, by Theorem 4.5. Optimal general and inner smoothings will then follow by (4.10). Now, observe that

$$f_\sigma^{\text{out}} = \left[r_\sigma + \sigma(\cdot - x_\sigma) \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \right] - w_\sigma = \left[\sigma(\cdot - x_\sigma) \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \right] - \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2$$

where the first equality is by definition and the second follows because $(c + f) \square g = r + (f \square g)$ for any $r \in \mathbb{R}$ and two functions f and g , and $w_\sigma = r_\sigma + \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2$ by definition. Therefore, $f_\sigma^{\text{out}} = \sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2$ follows immediately once we show that $x_\sigma = 0$. Further, from the identity $x_\sigma = 0$, we will get that $w_\sigma = r_\sigma = \varrho_\sigma(0) = \max_{\zeta \in \partial\sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2$ by Equations (4.1) and (4.2). Therefore proving that $x_\sigma = 0$ will conclude the proof.

Since $x_\sigma = \operatorname{argmin}_{x \in \mathcal{E}} \varrho_\sigma(x) + \frac{1}{2} \|x\|_\mathcal{E}^2$ by Equation 4.2, the first order condition $0 \in \partial\varrho_\sigma(0)$ is sufficient to prove $x_\sigma = 0$. Observe that (i) $\operatorname{argmax}_{\zeta \in \partial\sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2 \subseteq \partial\varrho_\sigma(0)$ by (4.1) and (ii) $\partial\sigma(0) = -\partial\sigma(0)$ because σ is a norm. Combining (i) and (ii) gives $0 = (\hat{\zeta} - \hat{\zeta})/2 \in \partial\varrho_\sigma(0)$ whenever $\hat{\zeta} \in \operatorname{argmax}_{\zeta \in \partial\sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2$, by convexity of the subdifferential. Since $\partial\sigma(0)$ is compact, it follows $\operatorname{argmax}_{\zeta \in \partial\sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2$ is nonempty. Therefore $0 \in \partial\varrho_\sigma(0)$ and the proof is complete. \square

Note that this result does not ensure that the Moreau envelope is the unique optimal outer smoothing. Depending on the structure of the norm under consideration, the optimal smoothing may or may not be unique. Below, we explicitly calculate two examples (one of each type) via their functional core, center, and width.

First consider the one-norm $\sigma(x) = \|x\|_1 := \sum_{i=1}^d |x_i|$ defined in $\mathcal{E} = \mathbb{R}^d$. Noting that $\partial\sigma(0) = \operatorname{conv}\{(\pm 1, \dots, \pm 1)\}$, we can calculate the functional core as

$$\varrho_\sigma(x) = \max_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_2^2 = \max_{\zeta \in \{(\pm 1, \dots, \pm 1)\}} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_2^2 = \sigma(x) + \frac{d}{2}.$$

Moreover, the functional center is given by $x_\sigma = (0, \dots, 0)$ and $r_\sigma = \frac{d}{2}$. Hence $\varrho_\sigma = r_\sigma + \sigma(\cdot - x_\sigma)$ and so (4.12) ensures the one-norm has a unique optimal smoothing. Proposition 4.8 then gives the unique optimal outer 1-smoothing of

$$f_\sigma^{\text{out}} = \|\cdot\|_1 \square \frac{1}{2} \|\cdot\|_2^2 = \sum_{i=1}^d H_1(x_i) \quad \text{where} \quad H_1(x_i) = \begin{cases} \frac{1}{2} |x_i|^2 & \text{if } |x_i| \leq 1 \\ |x_i| - \frac{1}{2} & \text{otherwise.} \end{cases}$$

The error of approximation is $\operatorname{dist}(f_\sigma^{\text{out}}, \sigma) = \frac{1}{2} w_\sigma = \frac{d}{2}$. The optimal general and inner 1-smoothings are then given by $\sum_{i=1}^d H_1(x_i) + \frac{d}{4}$ and $\sum_{i=1}^d H_1(x_i) + \frac{d}{2}$, respectively.

Second, as a less structured example norm, consider the following weighted infinity norm $\sigma(x) = \max\{|x_1|, 2|x_2|\}$ defined in $\mathcal{E} = \mathbb{R}^2$. Noting that $\partial\sigma(0) = \text{conv}\{\pm(1,0), \pm(0,2)\}$, one can calculate the functional core as

$$\varrho_\sigma(x) = \max_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_2^2 = \max_{\zeta \in \{\pm(1,0), \pm(0,2)\}} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_2^2 = \max\left\{|x_1| + \frac{1}{2}, 2|x_2| + 2\right\}.$$

Moreover, the functional center is given by $x_\sigma = (0,0)$ and $r_\sigma = 2$. Since $\sigma(x) + r_\sigma > \varrho_\sigma(x)$ whenever $|x_1| \geq 2$ and $|x_2| = 0$, we note that σ does not have unique optimal smoothings, by (4.12). In this case, the largest (in function value) optimal outer 1-smoothing is given by

$$f_\sigma^{\text{out}}(x) = \sigma \square \frac{1}{2} \|\cdot\|_2^2 = \begin{cases} \frac{1}{2} \|x\|_2^2 & \text{if } |x_1| + \frac{1}{2}|x_2| \leq 1 \\ |x_1| - \frac{1}{2} & \text{else if } 2|x_2| - |x_1| \leq -1 \\ 2|x_2| - 2 & \text{else if } 2|x_2| - |x_1| \geq 4 \\ \frac{1}{5} \left(4|x_1| + 2|x_2| - 2 + \frac{1}{2}(2|x_2| - |x_1|)^2\right) & \text{else} \end{cases}$$

while the smallest optimal outer 1-smoothing is given by

$$F_\sigma^{\text{out}}(x) = (\varrho_\sigma - w_\sigma) \square \frac{1}{2} \|\cdot\|_2^2 = \begin{cases} f_\sigma^{\text{out}}((0, x_2)) & \text{if } |x_1| \leq 3/2 \\ f_\sigma^{\text{out}}(x + (3/2, 0)) & \text{if } x_1 < -3/2 \\ f_\sigma^{\text{out}}(x - (3/2, 0)) & \text{else.} \end{cases}$$

Hence, a 1-smooth convex function f is an optimal 1-smoothing of this weighted infinity norm exactly if it lies between these two functions.

4.2.3 The Optimal Smoothings of Maximum and Maximum Eigenvalue Functions

Consider the max function $\sigma(x) := \max\{x_1, \dots, x_d\}$, defined in $\mathcal{E} = \mathbb{R}^d$. Noting that $\partial\sigma(0) = \text{conv}\{e_1, \dots, e_d\}$, where e_i is the i -th standard basis vector, we have

$$\begin{aligned} \varrho_\sigma(x) &= \max_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_2^2 = \max_{\zeta \in \{e_1, \dots, e_d\}} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_2^2 = \sigma(x) + \frac{1}{2}, \\ x_\sigma &= \operatorname{argmin}_{x \in \mathbb{R}^d} \frac{1}{2} + \sigma(x) + \frac{1}{2} \|x\|_2^2 = -(1, \dots, 1)/d, \\ r_\sigma &= \varrho_\sigma(x_\sigma) = -\frac{1}{d} + \frac{1}{2}. \end{aligned}$$

This implies $\varrho_\sigma = r_\sigma + \sigma(\cdot - x_\sigma)$ and so σ has a unique optimal 1-smoothing. In particular, it is given by³

$$f_\sigma^{\text{gen}}(x) = F_\sigma^{\text{gen}}(x) = \alpha + \frac{1}{2} \sum_{i=1}^J (x_{\pi(i)} - \alpha)^2 + \frac{1}{4} \left(1 + \frac{1}{d}\right) \quad (4.17)$$

where the permutation π ensures $x_{\pi(i)} \geq x_{\pi(i+1)}$ for all $i < d$, $\alpha = \frac{1}{J}(\sum_{i=1}^J x_{\pi(i)} - 1)$, and $J = \max\left\{j \in \{1, \dots, d\} \mid x_{\pi(j)} - \frac{1}{j}(\sum_{i=1}^j x_{\pi(i)} - 1) > 0\right\}$. The error of approximation is $\text{dist}(f_\sigma^{\text{gen}}, \sigma) = \frac{1}{2} w_\sigma = \frac{1}{4} \left(1 - \frac{1}{d}\right)$. More generally, $\frac{1}{\beta} f_\sigma^{\text{gen}}(\beta x)$ gives the optimal β -smoothing for any $\beta > 0$, with approximation error of $\frac{1}{4\beta} \left(1 - \frac{1}{d}\right)$.

Entirely similar reasoning and formulas apply to give the unique optimal smoothing of the maximum eigenvalue function $\sigma(A) := \max_{x \in \mathbb{R}^d, \|x\|_2=1} x^T A x$, defined in the space of symmetric $d \times d$ matrices $\mathcal{E} = \mathbb{S}^{d \times d}$. Appendix B provides the resulting parallel formulas.

³This can be verified by directly computing $F_\sigma^{\text{gen}}(x) = \varrho_\sigma \square \frac{1}{2} \|\cdot\|_2^2 - w_\sigma/2 = \left(\frac{1}{2} + \sigma\right) \square \frac{1}{2} \|\cdot\|_2^2 - (1 - 1/d)/4$ using (4.15) and the formula for projecting on the unit simplex $\partial\sigma(0) = \text{conv}\{e_1, \dots, e_d\}$ given in [29, Figure 1]. Further note from this projection formula, this optimal smoothing can be computed in (expected) linear time in n [30].

K	C_K	x_K	w_K	Unique?
\mathbb{R}_+^d	$\{(x_1, \dots, x_d) \in \mathbb{R}^d \mid x_i \geq 1, i = 1, \dots, d\}$	$(1, \dots, 1)$	$\sqrt{d} - 1$	Yes
K^{SOC}	$\{(x, t) \in \mathcal{E} \times \mathbb{R} \mid t - \sqrt{2} \geq \ x\ _2\}$	$(0, \sqrt{2})$	$\sqrt{2} - 1$	Yes
K^{SDP}	$\{x \in \mathbb{S}^{d \times d} \mid x \succeq I_d\}$	I_d	$\sqrt{d} - 1$	Yes
K^{EXP}	See Figure 2(c)	$(-1.11957, 1, 1.71471)$	1.27897	No

Table 1: Classic cones with their conic core, center and width. Values for the exponential cone are numerically computed. The final column indicates whether there is a unique optimal smoothing.

4.2.4 The Optimal Smoothings of Nonnegative, Second-Order, and Semidefinite Cone

Perhaps the three most common cones are the nonnegative orthant $\mathbb{R}_+^d = \{x \in \mathbb{R}^d \mid x \geq 0\}$, second-order cone $K^{\text{SOC}} = \{(x, t) \in \mathbb{R}^{d+1} \mid \|x\|_2 \leq t\}$, and positive semidefinite cone $K^{\text{SDP}} = \{X \in \mathbb{S}^{d \times d} \mid X \succeq 0\}$. Optimization over these sets corresponds to linear, second-order cone, and semidefinite programming.

Table 1 states the conic center and width for each of these cones. In each case, the conic core is exactly $x_K + K$, establishing that they each possess a unique optimal smoothing.

To illustrate the process of calculating these quantities, consider the nonnegative orthant $K = \mathbb{R}_+^d := \{x \in \mathbb{R}^d \mid x_i \geq 0, i = 1, \dots, d\}$. Then, $N_K(0) = \text{cone}\{-e_1, \dots, -e_d\}$ where e_i denotes the i -th standard basis vector. Letting $e = (1, \dots, 1)$ denote the all ones vector, we have that

$$C_K = \{x \in \mathcal{E} \mid \langle \zeta, x \rangle \leq -1, \|\zeta\|_2 = 1, \zeta \in N_K(0)\} = \{x \in \mathbb{R}^d \mid \langle -e_i, x \rangle \leq -1, i = 1, \dots, d\} = e + K,$$

$$x_K = \underset{x \in C_K}{\operatorname{argmin}} \|x\|_2 = e,$$

$$w_K = \|x_K\|_2 - 1 = \sqrt{d} - 1.$$

From this, it follows, for example, that $e + \mathbb{R}_+^d + B(0, 1)$ is the optimal inner 1-smoothing with approximation error $\sqrt{d} - 1$. Similar processes of calculating give the claimed quantities for the second-order and semidefinite cones.

4.2.5 Numerically Computing All Optimal Smoothings of the Exponential Cone Finally, providing a contrast to the previous three particularly nice cones, we consider the exponential cone

$$K^{\text{EXP}} = \{(x, y, z) \in \mathbb{R}^3 \mid z \geq ye^{x/y}, y > 0\} \cup \{(x, y, z) \in \mathbb{R}^3 \mid x \leq 0, z \geq 0, y = 0\}.$$

This cone has found widespread use, providing representations for many applied optimization problems in a standardized conic form. See the numerous references and examples in [31].

To the best of our knowledge, the conic core of this cone does not possess a closed form. Instead of an analytic solution, our definitions facilitate a numerical computation of the conic core, center, and width for this set. A rendering of the numerically computed conic core is given in Figure 2. An immediate observation from this is that the conic core is not a translated copy of the exponential cone, implying by (4.13) that there is not a unique optimal way to smooth the exponential cone. The set of all optimal inner 1-smoothings is characterized by the minimal and maximal optimal 1-smoothing s_K^{in} and S_K^{in} , respectively, as defined in Equations (4.6). Numerical renderings of these extremal smoothings are also shown in Figure 2.

We numerically estimate the boundary of the core C_K by sampling points x satisfying

$$\sup_{\zeta \in N_K(0), \|\zeta\|_{\mathcal{E}}=1} \langle \zeta, x \rangle = -1$$

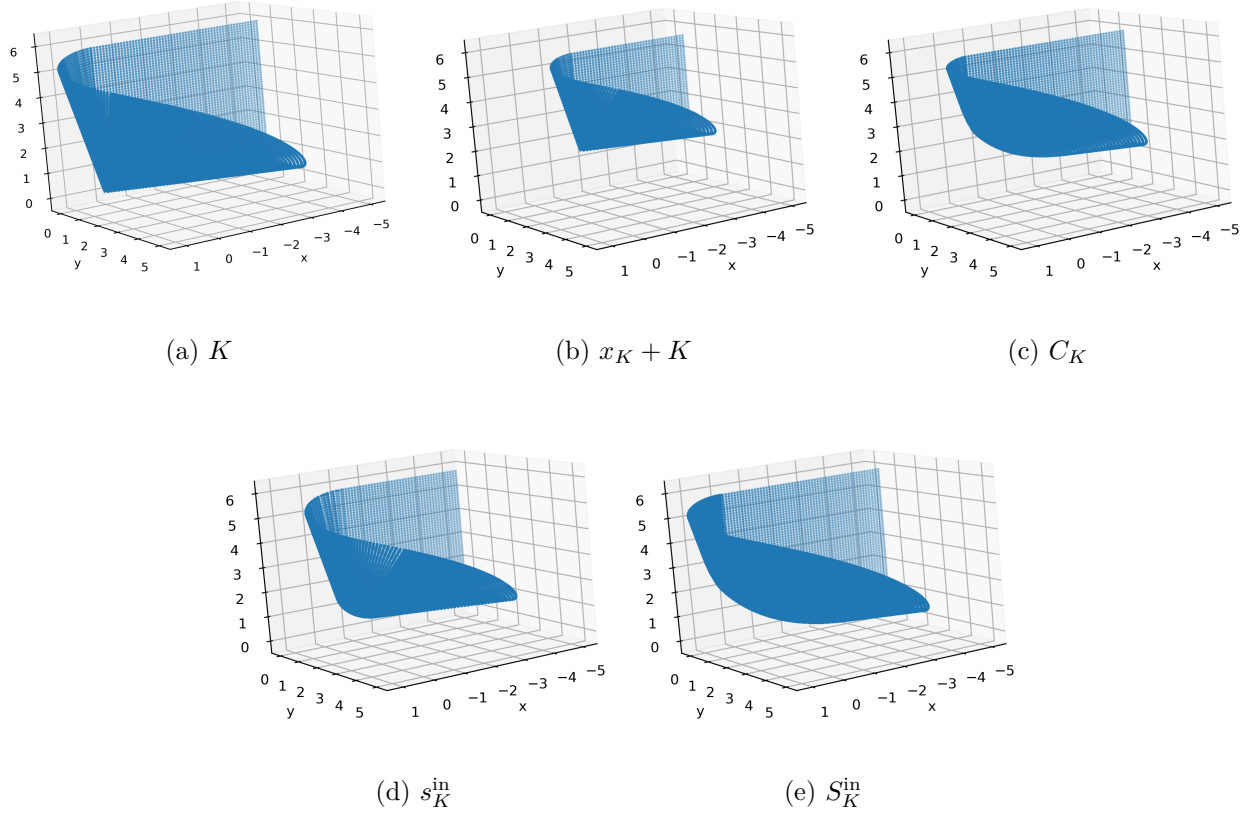


Figure 2: The exponential cone, its translation $x_K + K$, its conic core, its minimal optimal 1-smoothing, and its maximal optimal 1-smoothing.

as Lemma 4.4 dictates. The supremum is numerically computed using a dense sample from the set $\{\zeta \in \mathbb{R}^3 \mid \zeta \in N_K(0), \|\zeta\|_2 = 1\}$. By computing a projection onto this computed conic core, one can numerically estimate the exponential cone's conic center and width. These numerically computed values are given in Table 1. The smooth sets s_K^{in} and S_K^{in} are obtained by adding a ball to the points in $x_K + K$ and C_K , respectively.

4.3 Proofs of Characterizations of Optimal Smoothings

We begin with a simple technical lemma (Lemma 4.9). Then we prove our main theorems for sublinear function and convex cone smoothings, which follow the same pattern: (i) we first prove the result for inner β -smoothings, and then (ii) we show that characterizations of the optimal general and outer smoothings follow from simple transformations.

Lemma 4.9. *Let K be a nonempty closed convex cone and C be a nonempty closed convex set. Then $K = C^\infty$ whenever $\text{dist}(C, K) < \infty$.*

Proof. Let $r = \text{dist}(C, K)$. We have $C^\infty \subseteq (K + B_r)^\infty = K \subseteq (C + B_r)^\infty = C^\infty$ where the inclusions follow because $C \subseteq K + B_r$ and $K \subseteq C + B_r$, while the equalities follow by Proposition A.1. \square

4.3.1 Proof of Theorem 4.3 Our proof is broken into several lemmas which combine together to immediately yield the theorem. Recall, due to Lemma 3.3, it suffices to consider only $\beta = 1$. We

begin by showing in Lemma 4.10 that F_σ^{in} is an inner 1-smoothing that lower bounds any other inner 1-smoothing of σ .

Lemma 4.10. *F_σ^{in} is an inner 1-smoothing of σ with $\text{dist}(F_\sigma^{\text{in}}, \sigma) \geq w_\sigma$. Further, $F_\sigma^{\text{in}} \leq f$ for any convex inner 1-smoothing f .*

Proof. $F_\sigma^{\text{in}} = \varrho_\sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2$ is 1-smooth by [19, Proposition 12.60]. The inequality $F_\sigma^{\text{in}} \geq \sigma$ follows because

$$\text{epi } F_\sigma^{\text{in}} = \text{epi } \varrho_\sigma + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \subseteq \text{epi } \sigma$$

where the equality follows by (4.9) and the inclusion follows because $\text{epi } \varrho_\sigma = C_\sigma$ by Lemma 4.3. As a result, we have

$$\text{dist}(F_\sigma^{\text{in}}, \sigma) \geq F_\sigma^{\text{in}}(0) - \sigma(0) = w_\sigma$$

where the inequality holds because $F_\sigma^{\text{in}} \geq \sigma$ and the equality follows from Equation (4.2) and the definition of F_σ^{in} and the fact that $\sigma(0) = 0$. Now let $f : \mathcal{E} \rightarrow \mathbb{R}$ be any 1-smooth convex function such that $f \geq \sigma$. Then $f = g \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2$ for some convex function g , by [19, Proposition 12.60]. Note that $\text{epi } g \subseteq C_\sigma$ as (4.9) ensures $\text{epi } g + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 = \text{epi } f \subseteq \text{epi } \sigma$. Since $C_\sigma = \text{epi } \varrho_\sigma$ by Lemma 4.3, it follows that $g \geq \varrho_\sigma$. Hence $f = g \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \geq \varrho_\sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 = F_\sigma^{\text{in}}$. \square

We next show in Lemma 4.11 that f_σ^{in} is an inner 1-smoothing with that upper bounds all inner 1-smoothings f satisfying $\text{dist}(f, \sigma) \leq \text{dist}(f_\sigma^{\text{in}}, \sigma)$. Since any optimal f satisfies $\text{dist}(f, \sigma) \leq \text{dist}(f_\sigma^{\text{in}}, \sigma)$ by definition, this proves that f_σ^{in} upper bounds optimal smoothings.

Lemma 4.11. *f_σ^{in} is an inner 1-smoothing of σ with $\text{dist}(f_\sigma^{\text{in}}, \sigma) \leq w_\sigma$. Further, $f_\sigma^{\text{in}} \geq f$ for any convex, inner 1-smoothing f with $\text{dist}(f, \sigma) \leq w_\sigma$.*

Proof. f_σ^{in} is 1-smooth by [19, Proposition 12.60]. The inequality $f_\sigma^{\text{in}} \geq \sigma$ follows because

$$\text{epi } f_\sigma^{\text{in}} = \text{epi } (r_\sigma + \sigma(\cdot - x_\sigma)) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 = (x_\sigma, r_\sigma) + \text{epi } \sigma + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \subseteq \text{epi } \sigma + \text{epi } \sigma = \text{epi } \sigma$$

where the first equality follows by (4.9), the inclusion follows because $(x_\sigma, r_\sigma) \in C_\sigma$ ensures $(x_\sigma, r_\sigma) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \subseteq \text{epi } \sigma$, and the last equality follows because $\text{epi } \sigma$ is a convex cone. Therefore

$$\begin{aligned} \text{dist}(f_\sigma^{\text{in}}, \sigma) &= \sup_{x \in \mathcal{E}} f_\sigma^{\text{in}}(x) - \sigma(x) \\ &= \sup_{x \in \mathcal{E}} r_\sigma + \frac{1}{2} \|x - x_\sigma\|_\mathcal{E}^2 - \frac{1}{2} \|x - x_\sigma - P_{\partial\sigma(0)}(x - x_\sigma)\|_\mathcal{E}^2 - \sigma(x) \\ &= \sup_{x \in \mathcal{E}} r_\sigma + \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2 - \frac{1}{2} \|x_\sigma + P_{\partial\sigma(0)}(x - x_\sigma)\|_\mathcal{E}^2 + \langle P_{\partial\sigma(0)}(x - x_\sigma), x \rangle - \sigma(x) \\ &\leq r_\sigma + \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2 \\ &= w_\sigma \end{aligned}$$

where the second equality follows by (4.14), the inequality follows in part because of the subgradient inequality $\sigma(x) \geq \sigma(0) + \langle P_{\partial\sigma(0)}(x - x_\sigma), x \rangle$, and the last equality holds by definition of w_σ .

For the second part of the Lemma, note that if $\text{dist}(f, \sigma) \leq w_\sigma$, then

$$\begin{aligned} \text{epi } f_\sigma^{\text{in}} &= \text{epi } (r_\sigma + \sigma(\cdot - x_\sigma)) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 = (x_\sigma, r_\sigma) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 + \text{epi } \sigma \\ &= (x_\sigma, r_\sigma) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 + (\text{epi } f)^\infty \end{aligned}$$

where the first equality follows by (4.9) and the last equality follows because Lemma 4.9 ensures that $\text{epi } \sigma = (\text{epi } f)^\infty$ as $\text{dist}(\text{epi } f, \text{epi } \sigma) \leq \text{dist}(f, \sigma) \leq w_\sigma$. Therefore, to prove that $\text{epi } f_\sigma^{\text{in}} \subseteq \text{epi } f$, it suffices to show that $(x_\sigma, r_\sigma) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \subseteq \text{epi } f$. We next show that this inclusion holds. Recall that 1-smoothness implies $f(x) \leq f(0) + \langle \nabla f(0), x \rangle + \frac{1}{2} \|x\|_\mathcal{E}^2$ for all $x \in \mathcal{E}$ by (2.3). Therefore

$$\left(-\nabla f(0), f(0) - \frac{1}{2} \|\nabla f(0)\|_\mathcal{E}^2 \right) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \subseteq \text{epi } f \subseteq \text{epi } \sigma \quad (4.18)$$

where the first inclusion is by the above quadratic bound while the second follows by assumption. Our desired inclusion holds because of (4.18) and the next calculation, which shows that $x_\sigma = -\nabla f(0)$ and $r_\sigma = f(0) - \frac{1}{2} \|\nabla f(0)\|_\mathcal{E}^2$. We have

$$r_\sigma + \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2 \leq f(0) - \frac{1}{2} \|\nabla f(0)\|_\mathcal{E}^2 + \frac{1}{2} \|\nabla f(0)\|_\mathcal{E}^2 \leq r_\sigma + \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2$$

where the first inequality holds because $(-\nabla f(0), f(0) - \frac{1}{2} \|\nabla f(0)\|_\mathcal{E}^2) \in C_\sigma$ by (4.18), while the second follows because $f(0) = |f(0) - \sigma(0)| \leq \text{dist}(f, \sigma) \leq w_\sigma$. Equation (4.2) then ensures $x_\sigma = -\nabla f(0)$ and $r_\sigma = f(0) - \frac{1}{2} \|\nabla f(0)\|_\mathcal{E}^2$ as claimed. \square

Now, by Lemmas 4.10 and 4.11, the necessity of the condition $F_\sigma^{\text{in}} \leq f \leq f_\sigma^{\text{in}}$ is proved. For sufficiency, we note that if f is a convex inner 1-smoothing satisfying $F_\sigma^{\text{in}} \leq f \leq f_\sigma^{\text{in}}$, then

$$w_\sigma \leq \text{dist}(F_\sigma^{\text{in}}, \sigma) \leq \text{dist}(f, \sigma) \leq \text{dist}(f_\sigma^{\text{in}}, \sigma) \leq w_\sigma$$

where the first inequality follows by Lemma 4.10, and the second and third inequalities follow by assumption, while the last inequality by Lemma 4.11. Thus $\text{dist}(f, \sigma) = \text{dist}(F_\sigma^{\text{in}}, \sigma)$, hence f is optimal as Lemma 4.10 ensures F_σ^{in} is optimal. This concludes the proof of the second statement of the theorem.

The first part of the theorem also follows directly from Lemmas 4.10 and 4.11. Note that σ is inner λ -smoothable for all $\lambda \geq w_\sigma$ because $[f_\sigma^{\text{in}}]_{1/\beta}$ is an inner (w_σ, β) -smoothing of σ for all $\beta > 0$, by Lemma 4.11 and Lemma 3.3. On the other hand, if σ is λ -smoothable, then $\lambda \geq w_\sigma$ because $w_\sigma \leq \text{dist}(f, \sigma) \leq \lambda$ for any $(\lambda, 1)$ -smoothing f , where the first inequality follows by Lemma 4.10.

4.3.2 Proof of Theorem 4.4 The proof of this theorem follows the same style as that of Theorem 4.3. As in that proof, we restrict to the case $\beta = 1$ without loss of generality and break the proof into two key lemmas. The first, Lemma 4.12, proves the inclusion $S \subseteq S_K^{\text{in}}$ for any convex inner 1-smoothing S . The second, Lemma 4.13, proves the inclusion $s_K^{\text{in}} \subseteq S$ for any optimal smoothing S .

Lemma 4.12. *S_K^{in} is an inner 1-smoothing of K with $\text{dist}(S_K^{\text{in}}, K) \geq w_K$. Further, $S \subseteq S_K^{\text{in}}$ for any closed convex inner 1-smoothing S .*

Proof. Clearly $S_K^{\text{in}} = C_K + B(0, 1) \subseteq K$ by definition of C_K . Since C_K is convex and closed, S_K^{in} is 1-smooth by [20, Proposition 3]. We also have

$$\text{dist}(S_K^{\text{in}}, K) \geq \|P_{S_K^{\text{in}}}(0)\|_\mathcal{E} = \left\| x_K - \frac{x_K}{\|x_K\|_\mathcal{E}} \right\|_\mathcal{E} = \|x_K\|_\mathcal{E} - 1 = w_K$$

where the first equality follows because $-x_K \in N_{S_K^{\text{in}}}\left(x_K - \frac{x_K}{\|x_K\|_\mathcal{E}}\right)$ since for all $x \in S_K^{\text{in}}$ it holds that

$$\begin{aligned} \left\langle -x_K, x - \left(x_K - \frac{x_K}{\|x_K\|_\mathcal{E}} \right) \right\rangle &= \langle -x_K, P_{C_K}(x) - x_K \rangle + \langle -x_K, x - P_{C_K}(x) \rangle - \|x_K\|_\mathcal{E} \\ &= \langle 0 - P_{C_K}(0), P_{C_K}(x) - P_{C_K}(0) \rangle + \langle -x_K, x - P_{C_K}(x) \rangle - \|x_K\|_\mathcal{E} \\ &\leq 0 \end{aligned}$$

by the Cauchy-Schwarz inequality and the fact that $S_K^{\text{in}} = C_K + B(0, 1)$ ensures $\text{dist}(S_K^{\text{in}}, C_K) = 1$. Now let $S \subseteq K$ be 1-smooth. Then there exists a closed convex set C such that $S = C + B(0, 1)$, by [20, Proposition 3]. Since $C + B(0, 1) = S \subseteq K$, it follows that $C \subseteq C_K$ by definition of C_K . Therefore $S = C + B(0, 1) \subseteq C_K + B(0, 1) = S_K^{\text{in}}$, and the proof is complete. \square

Lemma 4.13. s_K^{in} is an inner 1-smoothing of K with $\text{dist}(s_K^{\text{in}}, K) \leq w_K$. Further, $s_K^{\text{in}} \subseteq S$ for any closed, convex optimal inner 1-smoothing S .

To prove Lemma 4.13 we will use the following lemma which will also be useful in other subsequent proofs.

Lemma 4.14. $d(C; S + B(0, r)) = (d(C; S) - r)_+$ for any closed, convex sets C and S .

Proof of Lemma 4.14. This holds by direct calculation as

$$\begin{aligned} d(C; S + B(0, r)) &= \sup_{y \in C} \|y - P_S(y) - P_{B_r}(y - P_S(y))\|_{\mathcal{E}} \\ &= \sup_{y \in C} (\|y - P_S(y)\|_{\mathcal{E}} - r)_+ \\ &= (d(C; S) - r)_+ \end{aligned}$$

where the first equality holds by (2.1), the second uses the simple fact that $\|z - P_{B_r}(z)\|_{\mathcal{E}} = (\|z\|_{\mathcal{E}} - r)_+$ for all $z \in \mathcal{E}$, while the last equality holds by continuity of the function $\alpha \mapsto (\alpha)_+ = \max\{0, \alpha\}$. \square

Proof of Lemma 4.13. s_K^{in} is 1-smooth by [20, Proposition 3]. Noting that $x_K \in C_K$ by definition of x_K , we have

$$s_K^{\text{in}} = x_K + B(0, 1) + K \subseteq K + K = K$$

where the first equality follows by definition of s_K^{in} and the inclusion follows because $x_K \in C_K \subseteq K$. We also have

$$\text{dist}(s_K^{\text{in}}, K) = d(K; x_K + K + B(0, 1)) = (d(K; x_K + K) - 1)_+ \leq (\|x_K\|_{\mathcal{E}} - 1)_+ = w_K$$

where the first equality holds because $s_K^{\text{in}} = x_K + K + B(0, 1) \subseteq K$, the second equality holds by Lemma 4.14, the inequality holds because $\|y - P_{(x_K + K)}(y)\|_{\mathcal{E}} \leq \|y - (x_K + y)\|_{\mathcal{E}} = \|x_K\|_{\mathcal{E}}$ for all $y \in K$, while the last equality holds by definition of w_K and the fact that $\|x_K\|_{\mathcal{E}} \geq 1$ as $x_K + B(0, 1) \subseteq K \neq \mathcal{E}$.

Now suppose S is a closed convex optimal inner 1-smoothing. We will show that $B(x_K, 1) \subseteq S$. This will imply $s_K^{\text{in}} \subseteq S$ since

$$s_K^{\text{in}} = x_K + K + B(0, 1) = B(x_K, 1) + S^{\infty} \subseteq S$$

where the second equality holds by Lemma 4.9 as optimality of S ensures $\text{dist}(S, K) \leq \text{dist}(s_K^{\text{in}}, K) \leq w_K < \infty$. Let $\zeta \in N_S(P_S(0))$ have $\|\zeta\|_{\mathcal{E}} = 1$. Such ζ exists because $P_S(0) \in \text{bdry } S$ since $P_S(0) \in \text{int } S$ would imply $0 \in \text{int } S \subseteq \text{int } K$, which contradicts the assumption that $K \neq \mathcal{E}$. We have $B(P_S(0) - \zeta, 1) \subseteq S$ by 1-smoothness of S . Therefore $B(x_K, 1) \subseteq S$ will be proved once we show that $x_K = P_S(0) - \zeta$. Observe that $P_S(0) - \zeta \in C_K$ because

$$P_S(0) - \zeta + B(0, 1) = B(P_S(0) - \zeta, 1) \subseteq S \subseteq K$$

where the first inclusion follows by 1-smoothness of S . Therefore

$$\|x_K\|_{\mathcal{E}} \leq \|P_S(0) - \zeta\|_{\mathcal{E}} = \|P_S(0)\|_{\mathcal{E}} + 1 \leq \text{dist}(S, K) + 1 \leq w_K + 1 = \|x_K\|_{\mathcal{E}} \quad (4.19)$$

where the first inequality holds because $x_K = P_{C_K}(0)$ and $P_S(0) - \zeta \in C_K$, while the first equality follows because $-P_S(0) = 0 - P_S(0) = \|P_S(0)\|_{\mathcal{E}} \zeta$, and the last inequality follows because $\text{dist}(S, K) \leq \text{dist}(s_K^{\text{in}}, K)$ by optimality assumption. Now, (4.19) implies $\|P_S(0) - \zeta - 0\|_{\mathcal{E}} = \|x_K\|_{\mathcal{E}} = \|P_{C_K}(0) - 0\|_{\mathcal{E}}$. Therefore $P_S(0) - \zeta = x_K$ by uniqueness of orthogonal projections. This proves that $B(x_K, 1) \subseteq S$, thereby concluding the proof. \square

From Lemmas 4.12 and 4.13, we immediately get the necessity of the condition $s_K^{\text{in}} \subseteq S \subseteq S_K^{\text{in}}$ on optimal smoothings. For sufficiency, note that if S is a closed, convex inner 1-smoothing that satisfies $s_K^{\text{in}} \subseteq S \subseteq S_K^{\text{in}}$, then

$$w_K \leq \text{dist}(S_K^{\text{in}}, K) \leq \text{dist}(S, K) \leq \text{dist}(s_K^{\text{in}}, K) \leq w_K$$

where the first inequality follows by Lemma 4.12, the second and third inequalities hold because $\text{dist}(\cdot, K)$ is monotone decreasing on subsets of K due to $\text{dist}(\cdot, K) = d(K; \cdot)$, while the last inequality holds by Lemma 4.13. Thus $\text{dist}(S, K) = \text{dist}(S_K^{\text{in}}, K)$, hence S is optimal as Lemma 4.12 ensures S_K^{in} is optimal. This proves the second statement of the theorem.

The first statement of Theorem 4.4 also follows immediately from Lemmas 4.12 and 4.13. Note K is inner λ -smoothable for all $\lambda \geq w_K$ because the set $[s_K^{\text{in}}]_{1/\beta}$ is an inner (w_K, β) -smoothing, by Lemmas 4.13 and 3.3. On the other hand, if K is inner λ -smoothable, then $\lambda \geq w_K$ as any inner $(\lambda, 1)$ -smoothing S of K has $w_K \leq \text{dist}(S, K) \leq \lambda$ where the first inequality follows from Lemma 4.12.

4.3.3 Proof of Theorems 4.1 and 4.5 Without loss of generality, we assume $\beta = 1$. Two key claims are needed for these proofs: for any convex, 1-smooth function f and any $r \geq 0$,

(A) If $f \geq \sigma$, then $f - r$ is 1-smooth and $\text{dist}(f - r, \sigma) \leq \max\{r, (\text{dist}(f, \sigma) - r)_+\}$.

(B) If $\text{dist}(f, \sigma) \leq r$, then $f = f^{\text{in}} - r$ where f^{in} is an inner 1-smoothing with

$$\text{dist}(f^{\text{in}}, \sigma) \leq \text{dist}((f - \sigma)_+, 0) + r.$$

Before proving these statements, we show that they immediately yield the theorems. Since $f^{\text{in}} \geq \sigma$ and $\text{dist}(f^{\text{in}}, \sigma) \leq w_\sigma$ for any optimal inner 1-smoothing by Lemma 4.11, (A) ensures that σ is $\max\{r, (w_\sigma - r)_+\}$ -smoothable with the smoothing $f_\sigma^{\text{in}} - r$. Further, for $r \geq w_\sigma$, this smoothing is an outer smoothing. Now, by (B), for any 1-smoothing f , we have $f = f^{\text{in}} - \text{dist}(f, \sigma)$ and

$$w_\sigma \leq \text{dist}(f^{\text{in}}, \sigma) \leq \text{dist}((f - \sigma)_+, 0) + \text{dist}(f, \sigma) \leq 2 \cdot \text{dist}(f, \sigma)$$

where the first inequality follows by Lemma 4.12, the second holds by (B), while the third holds because $\text{dist}(f, \sigma) = \max\{\text{dist}((f - \sigma)_+, 0), \text{dist}((\sigma - f)_+, 0)\}$. Similarly, noting that $(f - \sigma)_+ = 0$ if f is an outer smoothing, (B) ensures $\text{dist}(f, \sigma) \geq w_\sigma$ for any outer 1-smoothing f . Since $\max\{r, (w_\sigma - r)_+\} = r$ for $r \in \{\frac{w_\sigma}{2}, w_\sigma\}$, we have therefore shown that with $r = \frac{w_\sigma}{2}$ and $r = w_\sigma$, the transformation $f^{\text{in}} \mapsto f^{\text{in}} - r$ is a surjective map from optimal inner 1-smoothings to optimal general and outer 1-smoothings, respectively. Combining this observation with Theorem 4.3 and Equations (4.10) proves the if and only if characterization of the optimal general and outer smoothings.

We complete the proof by showing that (A) and (B) are true. For (A), clearly $f - r$ is as smooth as f . Further, if $f \geq \sigma$ then

$$\text{dist}(f - r, \sigma) = \sup_{x \in \mathcal{E}} \max \{r + \sigma(x) - f(x), f(x) - \sigma(x) - r\} \leq \max \{r, (\text{dist}(f, \sigma) - r)_+\}$$

where the inequality uses that $f(x) \geq \sigma(x)$ by assumption and $\text{dist}(f, \sigma) \geq f(x) - \sigma(x)$ for all $x \in \mathcal{E}$. For (B), if f is 1-smooth and $\text{dist}(f, \sigma) \leq r$, then clearly $f^{\text{in}} = f + r$ is 1-smooth and $f^{\text{in}} \geq \sigma$. Further

$$\text{dist}(f^{\text{in}}, \sigma) = \sup_{x \in \mathcal{E}} (f(x) + r - \sigma(x))_+ \leq \sup_{x \in \mathcal{E}} (f(x) - \sigma(x))_+ + r = \text{dist}((f - \sigma)_+, 0) + r$$

where the first equality holds because $f^{\text{in}} = f + r \geq \sigma$ while inequality uses that $r \geq 0$.

4.3.4 Proof of Theorems 4.2 and 4.6 The proof mirrors the proof of Theorems 4.1 and 4.5. As in that proof, we assume $\beta = 1$ with no loss in generality. We then show that the set transformation $S \mapsto [S + B(0, r)]_{1/(1+r)}$ (which is analogous to $f \mapsto f - r$) is a surjective map from optimal inner 1-smoothings and to general and outer optimal 1-smoothings when $r = \frac{w_K}{2}$ and $r = w_K$, respectively. The key observations are again that for any closed, convex, 1-smooth set S and any $r \geq 0$,

(A) If $S \subseteq K$, then $[S + B(0, r)]_{1/(1+r)}$ is 1-smooth and

$$\text{dist}([S + B(0, r)]_{1/(1+r)}, K) \leq \frac{\max \{r, (\text{dist}(S, K) - r)_+\}}{1 + r}.$$

(B) If $\text{dist}(S, K) \leq \frac{r}{1+r}$, then $S = [S^{\text{in}} + B(0, r)]_{1/(1+r)}$, where S^{in} is an inner 1-smoothing of K with

$$\text{dist}(S^{\text{in}}, K) \leq (1 + r)d(K; S) + r.$$

Proving that these statements are true is sufficient, for entirely symmetric arguments as in the previous proof of Theorems 4.1 and 4.5: smoothability claims follows from (A) by transforming an optimal inner 1-smoothing using the appropriate choice of $r \in \{\frac{w_K}{2}, w_K\}$, while the if and only if characterizations follow because of Theorem 4.3, Equations (4.11), and the fact that (A) and (B) together guarantee that $S \mapsto [S + B(0, r)]_{1/(1+r)}$ is a surjective map from optimal inner 1-smoothings to optimal general (with $r = \frac{w_K}{2}$) and outer (with $r = w_K$) 1-smoothings.

For (A), note that $S + B(0, r)$ is $(1 + r)^{-1}$ -smooth by [20, Proposition 3]. Therefore, by Lemma 3.3, $[S + B(0, r)]_{1/(1+r)}$ is 1-smooth. Further

$$\begin{aligned} \text{dist}(S + B(0, r), K) &= \max \{d(S + B(0, r); K), d(K; S + B(0, r))\} \\ &= \max \{d(S + B(0, r); K), (\text{dist}(S, K) - r)_+\} \\ &\leq \max \{r, (\text{dist}(S, K) - r)_+\} \end{aligned}$$

where the second equality holds by Lemma 4.14 and the fact that $S \subseteq K$ ensures $d(K; S) = \text{dist}(S, K)$; the inequality follows because $d(S + B_r; K) \leq d(K + B_r; K) = r$ due to $S \subseteq K$. Lemma 3.3 then ensures $\text{dist}([S + B(0, r)]_{1/(1+r)}, K) \leq \frac{\max \{r, (\text{dist}(S, K) - r)_+\}}{1 + r}$.

We now prove (B). Suppose $\text{dist}(S, K) \leq \frac{r}{1+r}$. Then Lemma 3.3 ensures $(1 + r)S$ is $(1 + r)^{-1}$ -smooth and $\text{dist}((1 + r)S, K) \leq r$. Therefore, there exists a closed, convex C such that $(1 + r)S =$

$C + B(0, 1 + r)$, by [20, Proposition 3]. Let $S^{\text{in}} = C + B(0, 1)$. Clearly, $(1 + r)S = S^{\text{in}} + B(0, r)$. Further, S^{in} is 1-smooth by [20, Proposition 3]. Note that $S^{\text{in}} \subseteq K$ holds by the contrapositive since every $x \notin K$ has

$$\begin{aligned} d(x + B(0, r); K) &\geq \left\| x + r \frac{x - P_K(x)}{\|x - P_K(x)\|_{\mathcal{E}}} - P_K \left(x + r \frac{x - P_K(x)}{\|x - P_K(x)\|_{\mathcal{E}}} \right) \right\|_{\mathcal{E}} \\ &= r + \|x - P_K(x)\|_{\mathcal{E}} \\ &> d(S^{\text{in}} + B(0, r); K) \end{aligned}$$

where the equality follows because $P_K \left(x + r \frac{x - P_K(x)}{\|x - P_K(x)\|_{\mathcal{E}}} \right) = P_K(x)$. The claimed distance bound holds because

$$\begin{aligned} \text{dist}(S^{\text{in}}, K) &= d(K; S^{\text{in}}) - r + r \leq (d(K; S^{\text{in}}) - r)_+ + r = d(K; (1 + r)S) + r \\ &= (1 + r)d(K; S) + r \end{aligned}$$

where the first equality follows from $S^{\text{in}} \subseteq K$, the inequality because $\alpha \leq (\alpha)_+$ for all $\alpha \in \mathbb{R}$, the second equality follows by Lemma 4.14, while the last equality holds because K is invariant to positive rescalings.

5 Applications: Nonsmooth Functions and Sets

Our theory can be used to handle nonsmoothness in more general settings. In this section, we demonstrate how this can be done for nonsmooth composite functions and nonsmooth sets with a nonempty interior.

5.1 Beyond Sublinear Functions: Application to Amenable Functions

We consider the smoothing of convex functions g of the form $g = \sigma \circ G$, under the assumption that

$$\|G(x) - G(y)\|_{\mathcal{E}'} \leq M \|x - y\|_{\mathcal{E}} \quad \text{and} \quad \|J(x) - J(y)\|_{\mathcal{E} \rightarrow \mathcal{E}'} \leq L \|x - y\|_{\mathcal{E}} \quad \text{for all } x, y \in \mathcal{E}. \quad (5.1)$$

Here $J(\cdot)$ denotes the Jacobian of $G : \mathcal{E} \rightarrow \mathcal{E}'$ and $\|\cdot\|_{\mathcal{E} \rightarrow \mathcal{E}'}$ is the operator norm. Nonsmooth functions of this type have been studied widely in optimization [32–34] and analysis [19, 35]. Our class of functions is a special subclass of amenable functions of Rockafellar and Wets [19, Definition 10.23], where σ is not assumed to be sublinear or finite everywhere and G is only required to have a continuous Jacobian. Our class is also closely related to decomposable functions of Shapiro (see [35, Definition 1.1]).

This class includes many nonsmooth functions that arise naturally in optimization; for example, minimizing finite maximums $\max\{G_1(x), \dots, G_n(x)\}$, maximum eigenvalue optimization $\lambda_{\max}(G(x))$, and nonlinear regression with an arbitrary norm $\|G(x)\|$.

Using our theory, amenable functions satisfying (5.1) can be smoothed by replacing the sublinear component σ with any smoothing. Doing so optimally gives the following guarantee on the smoothability of such functions.

Theorem 5.1. *Consider any function $g = \sigma \circ G$ where σ is sublinear and M_{σ} -Lipschitz and $G : \mathcal{E} \rightarrow \mathcal{E}'$ satisfies (5.1). Then g is $(M^2 w_{\sigma}/2, M_{\sigma}L)$ -smoothable, given by the β -smoothing*

$$x \mapsto \frac{M^2}{\beta - M_{\sigma}L} f_{\sigma}^{\text{gen}} \left(\frac{\beta - M_{\sigma}L}{M^2} G(x) \right).$$

Proof. First, we observe that for any β' -smoothing f_σ of σ , $f = f_\sigma \circ G$ is a $M_\sigma L + M^2 \beta'$ -smooth approximation to g with $\text{dist}(f_\sigma \circ G, g) \leq \text{dist}(\sigma, f_\sigma)$. To see this, note that

$$\text{dist}(f_\sigma \circ G, g) = \sup_{x \in \mathcal{E}} |\sigma(G(x)) - f_\sigma(G(x))| \leq \text{dist}(\sigma, f_\sigma).$$

Letting $J(\cdot)^T$ denote the adjoint operator of $J(\cdot)$, which is also L -Lipschitz continuous, we have

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\|_{\mathcal{E}} &= \|J(x)^T \nabla f_\sigma(G(x)) - J(y)^T \nabla f_\sigma(G(y))\|_{\mathcal{E}} \\ &= \|[J(x)^T - J(y)^T] \nabla f_\sigma(G(x)) + J(y)^T [\nabla f_\sigma(G(x)) - \nabla f_\sigma(G(y))]\|_{\mathcal{E}} \\ &\leq \|\nabla f_\sigma(G(x))\|_{\mathcal{E}'} \|J(x)^T - J(y)^T\|_{\mathcal{E}' \rightarrow \mathcal{E}} + \|J(y)^T\|_{\mathcal{E}' \rightarrow \mathcal{E}} \|\nabla f_\sigma(G(x)) - \nabla f_\sigma(G(y))\|_{\mathcal{E}'} \\ &\leq (M_\sigma L + M^2 \beta') \|x - y\|_{\mathcal{E}} \end{aligned}$$

where the first equality follows by the chain rule, the first inequality uses the triangle inequality and the definition of operator norm while the last inequality follows because that $F, J(\cdot)^T$ and ∇f_σ are Lipschitz continuous. In particular, the M -Lipschitz continuity of G ensures that $\|J(y)^T\|_{\mathcal{E}' \rightarrow \mathcal{E}} \leq M$. Moreover, the M_σ -Lipschitz continuity of σ ensures that $\|\nabla f_\sigma(G(x))\|_{\mathcal{E}} \leq M_\sigma$ since $\text{dist}(\sigma, f_\sigma) < \infty$ implies $\text{epi } \sigma = (\text{epi } f_\sigma)^\infty$ by Proposition A.1 and consequently $\nabla f_\sigma(z) \in \partial \sigma(0)$ for all $z \in \mathcal{E}'$.

From this bound, for any target smoothness $\beta > M_\sigma L$, consider selecting f_σ as an optimal $\beta' = (\beta - M_\sigma L)/M^2$ -smoothing of σ . This choice suffices to yield an $f = f_\sigma \circ G$ proving the claimed smoothability of $g = \sigma \circ G$. \square

5.1.1 Example: Finite Maximums and Improving LogSumExp Smoothing An important application where smoothing enables accelerated convergence theory is the minimization of a finite maximum

$$\min_x \max\{G_1(x), \dots, G_n(x)\}.$$

Supposing each G_i is convex and that $G(x) = (G_1(x), \dots, G_n(x))$ is M -Lipschitz, a subgradient method could be applied, giving a convergence guarantee of the form $M^2 \|x_0 - x_\star\|_{\mathcal{E}}^2 / \varepsilon^2$, where x_0 denotes the initial iterate and x_\star is an optimal solution. Beck and Teboulle [8] considered replacing $\sigma(x) = \max\{y_1, \dots, y_n\}$ above by the smoothing $\eta f^{\text{exp}}(G(x)/\eta)$ for small $\eta > 0$, where

$$f^{\text{exp}}(y) = \log \left(\sum_{i=1}^n \exp(y_i) \right).$$

Note that f^{exp} is 1-smooth and $\text{dist}(\sigma, f^{\text{exp}}) = \log(n)$. Therefore, if the Jacobian of G is L -Lipchitz, Lemma 3.3 and Theorem 5.1 ensure that $\eta f^{\text{exp}}(G(x)/\eta)$ is $L + M^2/\eta$ -smooth with $\max_{x \in \mathcal{E}} |\sigma(G(x)) - \eta f^{\text{exp}}(G(x)/\eta)| \leq \eta \log(n)$. By choosing $\eta = \varepsilon/(2 \log(n))$ and minimizing the smoothed objective to accuracy $\varepsilon/2$ using an accelerated gradient method (see [36]), this yields an improved convergence guarantee of

$$\sqrt{2L\varepsilon + 4 \log(n)M^2} \cdot \frac{\|x_0 - x_\star\|_2}{\varepsilon}.$$

Note f^{exp} is not the optimal 1-smoothing of σ . Rather the optimal smoothing is f_σ^{gen} given by (4.17). Using this smoothing and applying an accelerated gradient method to minimize $\varepsilon f_\sigma^{\text{gen}}(G(x)/\varepsilon)$ to accuracy $\varepsilon/2$ gives an overall convergence guarantee of

$$\sqrt{2L\varepsilon + 2M^2} \cdot \frac{\|x_0 - x_\star\|_2}{\varepsilon}.$$

This is an improvement by a factor of $1/\sqrt{\log(n)}$ in iteration complexity of prior works using the f^{exp} smoothing. In terms of per-iteration computational complexity, these methods are comparable: At each step, both methods must compute a Jacobian of G . Although computing the optimal smoothing (4.17) and its gradient is somewhat more complex than those of the logSumExp, it can still be done in (expected) linear time in n [30].

We note that the above improvement is not universal. This is because f^{exp} is also 1-smooth with respect to infinity norm, i.e. $\|\nabla f^{\text{exp}}(x) - \nabla f^{\text{exp}}(y)\|_1 \leq \|x - y\|_\infty$. Therefore, in calculating the smoothness constant of $f^{\text{exp}} \circ G$, one could instead use the \tilde{M} -Lipschitz and \tilde{L} -smooth constants under the infinity norm on \mathbb{R}^n . Recall, our smoothings are only optimal with respect to the two-norm. Since $\|\cdot\|_\infty \leq \|\cdot\|_2$, it follows that $\tilde{M} \leq M$ and $\tilde{L} \leq L$. Using this potentially more refined smoothness constant for f^{exp} , our theory provides an improved rate whenever $M/\tilde{M} \leq \sqrt{2\log(n)}$.

To attain a universal improvement on the f^{exp} smoothing, new theory is needed for optimal smoothings under general norms and, in particular, the infinity norm. We leave open the question of designing optimal smoothings under norms that do not have an inner product.

5.2 Beyond Convex Cones: Application to General Convex Sets

It is well-known that for any convex set $C \subseteq \mathcal{E}$, there exists a cone $K \subseteq \mathcal{E}' := \mathcal{E} \times \mathbb{R}$ and $x_0 \in \mathcal{E}$ such that $C = \{x \in \mathcal{E} \mid (x - x_0, 1) \in K\}$. Therefore, by smoothing K optimally, our theory can be used to approximate generic convex sets C . The following Theorem 5.2 shows that K can be chosen to be $\left(1 - \frac{R}{\sqrt{1+R^2}}\right)$ -smoothable whenever $B(x_0, R) \subseteq C$. Doing so, our optimal smoothings for K strictly improve upon the trivial 1-smoothing of $C + B(0, 1/\beta)$, at the cost of growing the problem dimension by one. These smoothings may be useful in optimization, as recent works [16, 20, 21] have proposed algorithms that benefit from the smoothness of constraints.

Theorem 5.2. *Let $C \subseteq \mathcal{E}$ be a closed, convex set with $B(x_0, R) \subseteq C$ for some $R > 0$. Then*

$$C = \{x \in \mathcal{E} \mid (x - x_0, 1) \in K\}$$

where K is the closure of the cone $\{(x, r) \in \mathcal{E} \times \mathbb{R} \mid r > 0, x/r \in C - x_0\}$. Further, K is closed, convex, and outer $\left(1 - \frac{R}{\sqrt{1+R^2}}\right)$ -smoothable.

Proof. We only prove smoothability since the other statements are obvious. By Theorem 4.2, it suffices to show that $\frac{w_K}{1+w_K} \leq 1 - \frac{R}{\sqrt{1+R^2}}$. Let $\mathcal{E}' = \mathcal{E} \times \mathbb{R}$. Recalling that $w_K = \|P_{C_K}(0)\|_{\mathcal{E}'} - 1$, and noting that $\frac{w_K}{1+w_K} \leq 1 - \frac{R}{\sqrt{1+R^2}} \iff w_K \leq \frac{\sqrt{1+R^2}-R}{R}$, it suffices to show that there exists $(\bar{x}, \bar{r}) \in C_K$ such that $\|(\bar{x}, \bar{r})\|_{\mathcal{E}'} - 1 \leq \frac{\sqrt{1+R^2}-R}{R}$. Taking $(\bar{x}, \bar{r}) = (0, \sqrt{1+R^2}/R)$, we have

$$\|(\bar{x}, \bar{r})\|_{\mathcal{E}'} - 1 = \frac{\sqrt{1+R^2}}{R} - 1 = \frac{\sqrt{1+R^2} - R}{R}.$$

Further, we have

$$\begin{aligned}
(\bar{x}, \bar{r}) \in C_K &\iff (\bar{x}, \bar{r}) + B(0, 1) \subseteq K \\
&\iff \left[\|x\|_{\mathcal{E}}^2 + \left(r - \frac{\sqrt{1+R^2}}{R} \right)^2 \leq 1 \implies (x, r) \in K \right] \\
&\iff \left[\|x\|_{\mathcal{E}}^2 + \left(r - \frac{\sqrt{1+R^2}}{R} \right)^2 \leq 1 \implies r > 0 \text{ and } \left\| \frac{x}{r} \right\|_{\mathcal{E}} \leq R \right] \\
&\iff \left[\|x\|_{\mathcal{E}}^2 + \left(r - \sqrt{1+R^{-2}} \right)^2 \leq 1 \implies \|x\|_{\mathcal{E}}^2 \leq r^2 R^2 \right] \\
&\iff \left[\|x\|_{\mathcal{E}}^2 \leq r^2 R^2 - \left(r^2 R^2 + \left(r - \sqrt{1+R^{-2}} \right)^2 - 1 \right) \implies \|x\|_{\mathcal{E}}^2 \leq r^2 R^2 \right] \\
&\iff r^2 R^2 + \left(r - \sqrt{1+R^{-2}} \right)^2 - 1 \geq 0 \\
&\iff (1+R^2)r^2 - 2r \frac{\sqrt{1+R^2}}{R} + \frac{1}{R^2} \geq 0 \\
&\iff \left(r\sqrt{1+R^2} - \frac{1}{R} \right)^2 \geq 0
\end{aligned}$$

where the first line follows by definition of C_K , the third line follows by definition of K and the assumption that $B(0, R) \subseteq C - x_0$, and the fourth because $\|x\|_{\mathcal{E}}^2 + \left(r - \sqrt{1+R^{-2}} \right)^2 \leq 1$ clearly implies $r \geq \sqrt{1+R^{-2}} - 1 > 0$. Therefore, $(\bar{x}, \bar{r}) \in C_K$ and the proof is complete. \square

Acknowledgements. This work was supported in part by the Air Force Office of Scientific Research under award number FA9550-23-1-0531. Benjamin Grimmer was additionally supported as a fellow of the Alfred P. Sloan Foundation.

References

- [1] Jean-Jacques Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.
- [2] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [3] Damek Davis and Dmitriy Drusvyatskiy. Stochastic model-based minimization of weakly convex functions. *SIAM Journal on Optimization*, 29(1):207–239, 2019.
- [4] J-P Penot and Mireille L Bougeard. Approximation and decomposition properties of some classes of locally dc functions. *Mathematical Programming*, 41(1):195–227, 1988.
- [5] Jean-Michel Lasry and Pierre-Louis Lions. A remark on regularization in hilbert spaces. *Israel Journal of Mathematics*, 55(3):257–266, 1986.
- [6] Hedy Attouch and Dominique Aze. Approximation and regularization of arbitrary functions in hilbert spaces by the lasry-lions method. In *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*, volume 10, pages 289–312. Elsevier, 1993.
- [7] A. Ben-Tal and M. Teboulle. A smoothing technique for nondifferentiable optimization problems. In Szymon Dolecki, editor, *Optimization*, pages 1–11, Berlin, Heidelberg, 1989. Springer Berlin Heidelberg.

- [8] Amir Beck and Marc Teboulle. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22:557–580, 2012.
- [9] Jonah Sherman. Nearly maximum flows in nearly linear time. In *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, FOCS '13, page 263–269, USA, 2013. IEEE Computer Society.
- [10] Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.
- [11] Ming Lei, Ting Kei Pong, Shuqin Sun, and Man-Chung Yue. Subdifferentially polynomially bounded functions and gaussian smoothing-based zeroth-order optimization. *arXiv preprint arXiv:2405.04150v3*, 2024.
- [12] Jelena Diakonikolas and Cristóbal Guzmán. Optimization on a finer scale: Bounded local subgradient variation perspective. *arXiv preprint arXiv:2303.05037v2*, 2024.
- [13] Fritz John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays, Presented to R. Courant on his 60th Birthday January 8, 1948*, pages 187–204. Wiley Interscience, New York, 1948.
- [14] Leonid G Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.
- [15] Yurii Nesterov. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimization Methods and Software*, 23(1):109–142, 2008.
- [16] Alan Luner and Benjamin Grimmer. Performance estimation for smooth and strongly convex sets. *arXiv preprint arXiv:2410.14811v2*, 2024.
- [17] Robert Deville, Vladimir Fonf, and Petr Hájek. Analytic and polyhedral approximation of convex bodies in separable polyhedral banach spaces. *Israel Journal of Mathematics*, 105(1):139–154, 1998.
- [18] Jiefeng Xu, Ting Kei Pong, and Nung sing Sze. A smoothing moving balls approximation method for a class of conic-constrained difference-of-convex optimization problems. *arXiv preprint arXiv:2505.12314v1*, 2025.
- [19] Roger J. B. Wets R. Tyrrell Rockafellar. *Variational Analysis*. Springer Berlin, Heidelberg, 2009.
- [20] Ning Liu and Benjamin Grimmer. Gauges and accelerated optimization over smooth and/or strongly convex sets. *arXiv preprint arXiv:2303.05037v3*, 2024.
- [21] Thabo Samakhoana and Benjamin Grimmer. Scalable projection-free optimization methods via multiradial duality theory. *arXiv preprint arXiv:2403.13688v2*, 2024.
- [22] Francis H. Clarke, Yuri S. Ledyaev, Ronald J. Stern, and Peter R. Wolenski. *Nonsmooth Analysis and Control Theory*. Graduate Texts in Mathematics. Springer New York, NY, 1998.
- [23] Zikai Xiong and Robert M. Freund. The role of level-set geometry on the performance of pdhg for conic linear optimization. *arXiv preprint arXiv:2406.01942v3*, 2024.
- [24] Robert M. Freund and Jorge R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM Journal on Optimization*, 10(1):155–176, 1999.
- [25] Jonathan T. Barron. Squareplus: A softplus-like algebraic rectifier. *arXiv preprint arXiv:2112.11687v1*, 2021.
- [26] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941v2*, 2017.
- [27] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

- [28] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289v5*, 2016.
- [29] John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the l_1 -ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pages 272–279, 2008.
- [30] Laurent Condat. Fast projection onto the simplex and the l_1 ball. *Math. Program.*, 158(1–2):575–585, July 2016.
- [31] Stephen Boyd, Seung-Jean Kim, Lieven Vandenbergh, and Arash Hassibi. A tutorial on geometric programming. *Optimization and Engineering*, 8(1):67–127, 2007.
- [32] James V. Burke. Descent methods for composite nondifferentiable optimization problems. *Mathematical Programming*, 33(3):260–279, 1985.
- [33] Yurii Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.
- [34] A. S. Lewis and S. J. Wright. A proximal method for composite minimization. *Mathematical Programming*, 158(1):501–546, 2016.
- [35] Alexander Shapiro. On a class of nonsmooth composite functions. *Mathematics of Operations Research*, 28(4):677–692, 2003.
- [36] Yurii Nesterov. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

A Horizon Cone Invariant to Ball Addition

Proposition A.1. *Let C be a nonempty, closed, convex set and $r \geq 0$. Then $(B_r + C)^\infty = C^\infty$.*

Proof. Note that $C^\infty \subseteq (B_r + C)^\infty$ because $C \subseteq B_r + C$. Now let $d \in (B_r + C)^\infty$. It suffices to show that there exists $c \in C$ such that $c + d \in C$. For any $c, \bar{c} \in C$ and $\zeta \in N_C(\bar{c})$, we have

$$\begin{aligned} \langle \zeta, d \rangle &= \langle \zeta, c + d - P_C(c + d) \rangle + \langle \zeta, P_C(c + d) - \bar{c} \rangle + \langle \zeta, \bar{c} - c \rangle \\ &\leq \|\zeta\|_{\mathcal{E}} \|c + d - P_C(c + d)\|_{\mathcal{E}} + \langle \zeta, \bar{c} - c \rangle \\ &\leq r \|\zeta\|_{\mathcal{E}} + \langle \zeta, \bar{c} - c \rangle \end{aligned}$$

where the first inequality is by the Cauchy-Schwartz inequality and normality conditions, while the second follows because $c + d \in B_r + C$. Since d was arbitrary and $(B_r + C)^\infty$ is a cone, it follows that $\zeta \in N_{(B_r + C)^\infty}(0)$. Therefore, $\langle \zeta, c + d - \bar{c} \rangle \leq \langle \zeta, c - \bar{c} \rangle \leq 0$. We have thus shown that

$$c + d \in \{x \in \mathcal{E} \mid \langle \zeta, x - \bar{c} \rangle \leq 0, \bar{c} \in C, \zeta \in N_C(\bar{c})\} = C$$

where the equality follows because C is closed and convex. □

B Optimal Smoothing of the Maximum Eigenvalue Function

Consider $\mathcal{E} = \mathbb{S}^{d \times d}$, the space of symmetric $d \times d$ matrices with the trace inner product and Frobenious norm. Let $\sigma(A) := \lambda_{\max}(A) := \max_{x \in \mathbb{R}^d, \|x\|_2=1} x^T A x$. Note that $\partial\sigma(0) = \text{conv}\{xx^T \mid$

$x \in \mathbb{R}^d, \|x\|_2 = 1\}$. Therefore, the functional core and center are

$$\begin{aligned}\varrho_\sigma(A) &= \max_{\zeta \in \partial\sigma(0)} \langle \zeta, A \rangle + \frac{1}{2} \|\zeta\|_{\mathcal{E}}^2 = \max_{\zeta \in \{\zeta x^T \mid x \in \mathbb{R}^d, \|x\|_2 = 1\}} \langle \zeta, A \rangle + \frac{1}{2} = \sigma(A) + \frac{1}{2}, \\ x_\sigma &= \operatorname{argmin}_{A \in \mathbb{S}^{d \times d}} \frac{1}{2} + \sigma(A) + \frac{1}{2} \|A\|_{\mathcal{E}}^2 = -I_d/d, \\ r_\sigma &= -\frac{1}{d} + \frac{1}{2}\end{aligned}$$

where I_d is the identity matrix for \mathbb{R}^d . This implies $\varrho_\sigma = r_\sigma + \sigma(\cdot - x_\sigma)$ and so σ has a unique optimal 1-smoothing. In particular, it is given by⁴

$$f_{\lambda_{\max}}^{\text{gen}}(A) = \lambda_{\max} \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2(A) + \frac{1}{4} \left(1 + \frac{1}{d}\right) = \alpha + \frac{1}{2} \sum_{j=1}^J (\lambda_j(A) - \alpha)^2 + \frac{1}{4} \left(1 + \frac{1}{d}\right)$$

where $\lambda_1(A), \dots, \lambda_d(A)$ are the eigenvalues of A sorted in descending order, $\alpha = \frac{1}{J}(\sum_{i=1}^J \lambda_i(A) - 1)$, and $J = \max \left\{j \in \{1, \dots, d\} \mid \lambda_j(A) - \frac{1}{j}(\sum_{i=1}^j \lambda_i(A) - 1) > 0\right\}$.

The error of approximation is $\max_{A \in \mathbb{S}^{d \times d}} \left| f_{\lambda_{\max}}^{\text{gen}}(A) - \sigma(A) \right| = \frac{1}{2} w_\sigma = \frac{1}{4} (1 - \frac{1}{d})$. More generally, $\frac{1}{\beta} f_{\lambda_{\max}}^{\text{gen}}(\beta A)$ gives the optimal β -smoothing for any $\beta > 0$, with approximation error of $\frac{1}{4} (1 - \frac{1}{d}) / \beta$.

⁴This can be verified by directly computing $\varrho_\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 - w_\sigma/2 = (\frac{1}{2} + \sigma) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 - (1 - 1/d)/4$ using Equation (4.15) and noting that $P_{\partial\sigma(0)}(A) = \sum_{i=1}^d (\lambda_i(A) - \alpha)_+ v_i v_i^T$ if $A = \sum_{i=1}^d \lambda_i(A) v_i v_i^T$ is the spectral decomposition of A .