

The Optimal Smoothings of Sublinear Functions and Convex Cones

Thabo Samakhoana^{1*} and Benjamin Grimmer¹

¹Applied Mathematics and Statistics, Johns Hopkins University, 3400
N. Charles Street, Baltimore, 21218, Maryland, USA.

*Corresponding author(s). E-mail(s): tsamakh1@jhu.edu;
Contributing authors: grimmer@jhu.edu;

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. Our constructed smoothings, by virtue of their optimality, establish a fundamental limit on potential applications of smoothing techniques. 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.

Keywords: Convex cones, Optimal, Smoothing, Sublinear functions

MSC Classification: 90C25 , 90C30

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 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:

- 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. Although these approaches have found practical success, they leave open whether yet stronger smoothing choices are possible.

Our Contributions.

This 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. By proving optimality, we establish fundamental limits on the capabilities of the many above-mentioned applications of smoothings. 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 (cf. Beck and Teboulle [8, Theorem 3.1, Example 4.9, and Proposition 4.1], adapted to our smoothness assumptions on G). Our theory provides an optimal smoothing f_σ^{gen} (see Equation (23)) 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 19 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 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].

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 6.1.

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 about sublinear functions are stated and proven in Section 4, characterizing the set of all optimal smoothings of any sublinear function. Section 5 mirrors Section 4 by characterizing the set of all optimal smoothings of any convex cone. Finally, Section 6 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 $\mathcal{E}^*$ 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 $\text{conv } C$.

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].

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

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

We denote the orthogonal projection onto C by $P_C(x) = \operatorname{argmin}\{\|x - y\|_{\mathcal{E}} \mid y \in C\}$. 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 for any closed C , $B_r + C = \{x + u \mid \|u\|_{\mathcal{E}} \leq r, x \in C\} = \{y \in \mathcal{E} \mid \operatorname{dist}(y, C) \leq r\}$. Note that if C is closed and convex, then

$$P_{B(x,r)+C}(y) = x + P_C(y - x) + P_{B_r}(y - x - P_C(y - x)) \quad (1)$$

for all $x, y \in \mathcal{E}$ and $r \geq 0$. To prove this, let $(\alpha)_+$ denote $\max\{0, \alpha\}$ from now on, where $\alpha \in \mathbb{R}$. Then for $u = y - x - P_C(y - x)$, $w \in B(x, r)$, and $z \in C$, we have

$$\begin{aligned} & \|u\|_{\mathcal{E}} \langle u - P_{B_r}(u), w + z - x - P_C(y - x) - P_{B_r}(u) \rangle \\ &= \|u\|_{\mathcal{E}} \langle u - P_{B_r}(u), w - x - P_{B_r}(u) \rangle + (\|u\|_{\mathcal{E}} - r)_+ \langle u, z - P_C(y - x) \rangle \\ &\leq 0 \end{aligned}$$

where the equality holds because $\|v\|_{\mathcal{E}} (v - P_{B_r}(v)) = (\|v\|_{\mathcal{E}} - r)_+ v$ for all $v \in \mathcal{E}$, while the inequality holds by normality conditions for the sets B_r and C . Plugging in the definition of u in the above inequality shows that $x + P_C(y - x) + P_{B_r}(y - x - P_C(y - x))$ satisfies the required normality conditions because $w + z$ is an arbitrary element of $B(x, r) + C$. This proves Equation (1).

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

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

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

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

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

$$\operatorname{dist}(f, g) := \sup_{x \in \mathcal{E}} |f(x) - g(x)|. \quad (2)$$

For such functions, we write $f \geq g$ to denote that $f(x) \geq g(x)$ for all $x \in \mathcal{E}$. A function $\sigma : \mathcal{E} \rightarrow (-\infty, \infty]$ is sublinear if $\sigma(x + y) \leq \sigma(x) + \sigma(y)$ and $\sigma(\lambda x) = \lambda \sigma(x)$ for all $x, y \in \mathcal{E}$ and $\lambda \geq 0$. For real-valued sublinear functions, the following useful identities hold (see [19, Theorem 8.13 and Corollary 8.25])

$$\sigma(x) = \sup_{\zeta \in \partial \sigma(0)} \langle \zeta, x \rangle \quad \text{and} \quad \partial \sigma(x) = \operatorname{argmax}_{\zeta \in \partial \sigma(0)} \langle \zeta, x \rangle \quad \text{for all } x \in \mathcal{E}. \quad (3)$$

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

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

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|y - x\|_{\mathcal{E}}^2 \quad (4)$$

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 (4) holds for a closed convex set C (see [20, Proposition 3]),

$$C \text{ is } \beta\text{-smooth} \iff B(x - \zeta/\beta, 1/\beta) \subseteq C \quad \text{for all unit norm } \zeta \in N_C(x). \quad (5)$$

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

$$f \text{ is } \beta\text{-smooth} \implies \text{epi } f \text{ is } \beta\text{-smooth} \quad (6)$$

for all convex $f : \mathcal{E} \rightarrow \mathbb{R}$. Note that the converse to (6) 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 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.

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 f to be globally smooth and (ii) 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 2 A closed 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

$$[f]_\eta(x) = \eta f(x/\eta) \quad [C]_\eta = \eta C = \{\eta x \mid x \in C\}. \quad (7)$$

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 1 *For any $\eta > 0$, function f and sublinear σ , and set C and cone K ,*

$$\text{dist}([f]_\eta, \sigma) = \eta \cdot \text{dist}(f, \sigma) \quad \text{and} \quad \text{dist}([C]_\eta, K) = \eta \cdot \text{dist}(C, K). \quad (8)$$

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

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

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 1, 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 1 and 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 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 4 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.

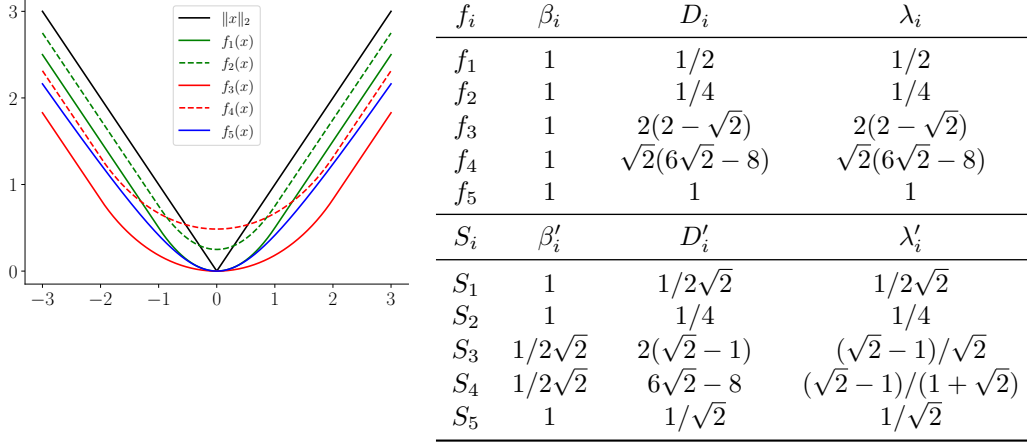


Fig. 1: Left: Five candidate smoothings of $\|x\|_2$ in 1D. Right: Smoothness and distance bounds, and the implied smoothability bound for the two-norm and second-order cone from each candidate.

For each $i = 1, 2, \dots, 5$, Figure 1 computes constants for the β_i -smoothness of each f_i and the distance $D_i = \text{dist}(f_i, \sigma)$.³ Since $\sigma = \|\cdot\|_2$ is sublinear, Lemma 1 ensures $[f_i]_\eta$ is β_i/η smoothing 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 3 and 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 1, we know $[S_i]_\eta$ is β'_i/η smoothing with distance from K of $\eta D'_i$. 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 12 and 14 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 Sections 4.3 and 5.3. 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.

³This can be verified for each function by noting in each case $D_i = \max\{f_i(0), \lim_{\|x\|_2 \rightarrow \infty} \sigma(x) - f_i(x)\}$.

4 Optimal Smoothings of Sublinear Functions

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

$$\sigma(\lambda x) = \lambda \sigma(x) \text{ if } \lambda \geq 0, \quad \sigma(x + y) \leq \sigma(x) + \sigma(y) \quad \text{for all } x \in \mathcal{E}.$$

Assumption (Blanket Assumption about σ) *For the rest of the paper, we assume σ is a real-valued sublinear function.*

Note that any sublinear function σ is β -smoothable for any $\beta > 0$ by, for example, considering the Moreau envelope of σ . Our focus here is not just demonstrating some smoothing f with $\text{dist}(\sigma, f) < \infty$, but rather in functions minimizing this gap.

Definition 5 A convex function $f : \mathcal{E} \rightarrow \mathbb{R}$ is an optimal β -smoothing of sublinear $\sigma : \mathcal{E} \rightarrow \mathbb{R}$ if f is β -smooth and $\text{dist}(\sigma, f) \leq \text{dist}(\sigma, \tilde{f})$ for all β -smooth, convex $\tilde{f} : \mathcal{E} \rightarrow \mathbb{R}$.

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

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

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

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

We focus on the epigraphical representation of σ in anticipation of the symmetry that exists between smoothing sublinear functions and convex cones. The functional core C_σ is the epigraph of the following closed, convex function:

$$\varrho_\sigma(x) := \max_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_{\mathcal{E}}^2. \quad (9)$$

Lemma 2 *For any sublinear function σ , it holds that $C_\sigma = \text{epi } \varrho_\sigma$.*

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

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

$$\iff (x, r) \in \text{epi } \varrho_\sigma$$

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

$$\begin{aligned} \sup_{y \in \mathcal{E}} \sigma(y) - \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 &= \sup_{y \in \mathcal{E}} \sup_{\zeta \in \partial\sigma(0)} \langle \zeta, y \rangle - \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 \\ &= \sup_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \sup_{y \in \mathcal{E}} \langle \zeta, y - x \rangle - \frac{1}{2} \|y - x\|_{\mathcal{E}}^2 \\ &= \sup_{\zeta \in \partial\sigma(0)} \langle \zeta, x \rangle + \frac{1}{2} \|\zeta\|_{\mathcal{E}}^2 \end{aligned}$$

where the first equality holds because sublinearity via (3) implies $\sigma(y) = \sup_{\zeta \in \partial\sigma(0)} \langle \zeta, y \rangle$ for all $y \in \mathcal{E}$. $\square$

From Lemma 2, it follows that $\min_{(x,r) \in C_\sigma} r + \frac{1}{2} \|x\|_{\mathcal{E}}^2$ has a unique solution given by

$$x_\sigma = \operatorname{argmin}_{x \in \mathcal{E}} \varrho_\sigma(x) + \frac{1}{2} \|x\|_{\mathcal{E}}^2 \quad \text{and} \quad r_\sigma = \varrho_\sigma(x_\sigma) \quad (10)$$

since

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

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

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:

$$f_\sigma^{\text{gen}} = \left[(r_\sigma + \sigma(\cdot - x_\sigma)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma/2, \quad F_\sigma^{\text{gen}} = \left[\varrho_\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma/2 \quad (11)$$

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.

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

The following theorem establishes that these are 1-smoothings and, in fact, provide the smallest and largest optimal 1-smoothings of σ , and suffice to characterize the set of all optimal smoothings.

Theorem 3 (Optimal Smoothings of σ) *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 $[F_\sigma^{\text{gen}}]_{1/\beta} \leq f \leq [f_\sigma^{\text{gen}}]_{1/\beta}$. Further, any such f has $\text{dist}(f, \sigma) = \frac{w_\sigma}{2\beta}$.*

We defer the proof of this theorem to Section 4.2.2 since it can be directly proved from the following Theorem 4 characterizing optimal inner smoothings.

4.1.1 Characterizations for Optimal Inner and Outer Smoothings

Definition 6 A convex function $f : \mathcal{E} \rightarrow \mathbb{R}$ is an optimal inner β -smoothing of sublinear $\sigma : \mathcal{E} \rightarrow \mathbb{R}$ if f is β -smooth, $f \geq \sigma$, and $\text{dist}(\sigma, f) \leq \text{dist}(\sigma, \tilde{f})$ for all β -smooth, convex $\tilde{f} : \mathcal{E} \rightarrow \mathbb{R}$ satisfying $\tilde{f} \geq \sigma$. Analogously, a β -smooth convex $f \leq \sigma$ is an optimal outer β -smoothing of σ if $\text{dist}(\sigma, f) \leq \text{dist}(\sigma, \tilde{f})$ for all β -smooth, convex $\tilde{f} \leq \sigma$.

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

$$f_\sigma^{\text{in}} = (r_\sigma + \sigma(\cdot - x_\sigma)) \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2, \quad F_\sigma^{\text{in}} = \varrho_\sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2. \quad (12)$$

Note that (see [19, Exercise 1.28])

$$\text{epi } f \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 = \text{epi } f + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \quad \text{for all convex } f : \mathcal{E} \rightarrow \mathbb{R} \quad (13)$$

and

$$\begin{aligned} \text{epi } r_\sigma + \sigma(\cdot - x_\sigma) &= \{(x, r) \mid r_\sigma + \sigma(x - x_\sigma) \leq r\} = \{(\tilde{x}, \tilde{r}) + (x_\sigma, r_\sigma) \mid \sigma(\tilde{x}) \leq \tilde{r}\} \\ &= (x_\sigma, r_\sigma) + \text{epi } \sigma \end{aligned}$$

via the change of variable $(\tilde{x}, \tilde{r}) = (x - x_\sigma, r - r_\sigma)$. Therefore, combining (12), (13), and the fact that $\text{epi } \varrho_\sigma = C_\sigma$ by Lemma 2, we get

$$\text{epi } f_\sigma^{\text{in}} = ((x_\sigma, r_\sigma) + \text{epi } \sigma) + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \quad \text{epi } F_\sigma^{\text{in}} = C_\sigma + \text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2. \quad (14)$$

Thus, the functions in (12) have a simple geometric interpretation: f_σ^{in} is obtained by adding $\text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2$ to a translation of $\text{epi } \sigma$ while F_σ^{in} is obtained by adding $\text{epi } \frac{1}{2} \|\cdot\|_\mathcal{E}^2$ to the functional core C_σ . Theorem 4 shows that these functions are optimal 1-smoothings and any other optimal 1-smoothing must lie between them.

Theorem 4 (Optimal Inner Smoothings of σ) *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 $[F_\sigma^{\text{in}}]_{1/\beta} \leq f \leq [f_\sigma^{\text{in}}]_{1/\beta}$. Further, any such f has $\text{dist}(f, \sigma) = \frac{w_\sigma}{\beta}$.*

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

$$f_\sigma^{\text{out}} = \left[(r_\sigma + \sigma(\cdot - x_\sigma)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma, \quad F_\sigma^{\text{out}} = \left[\varrho_\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right] - w_\sigma \quad (15)$$

Theorem 5 formalizes this and is proven in Section 4.2.2.

Theorem 5 (Optimal Outer Smoothings of σ) *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 $[F_\sigma^{\text{out}}]_{1/\beta} \leq f \leq [f_\sigma^{\text{out}}]_{1/\beta}$. Further, any such f has $\text{dist}(f, \sigma) = \frac{w_\sigma}{\beta}$.*

4.1.2 Two Remarks on Optimal Smoothings

First, we note that one can obtain general and outer smoothings by simply adding a constant to inner smoothings, as

$$\begin{aligned} f_\sigma^{\text{gen}} &= f_\sigma^{\text{in}} - w_\sigma/2, & F_\sigma^{\text{gen}} &= F_\sigma^{\text{in}} - w_\sigma/2, \\ f_\sigma^{\text{out}} &= f_\sigma^{\text{in}} - w_\sigma, & F_\sigma^{\text{out}} &= F_\sigma^{\text{in}} - w_\sigma. \end{aligned} \quad (16)$$

These reformulations can easily be verified from definitions.

Second, we note that our theory provides an exact characterization of when optimal smoothing of σ are unique:

$$\sigma \text{ has unique smoothings} \iff f_\sigma^{\text{in}} = F_\sigma^{\text{in}} \iff r_\sigma + \sigma(\cdot - x_\sigma) = \varrho_\sigma \quad (17)$$

where the first equivalence follows from our main theorems and the second uses the elementary fact that two closed convex functions are equal if and only if their Moreau envelopes are equal⁴.

4.2 Proofs of Characterizations of Optimal Smoothings

We begin with a simple technical lemma (Lemma 6) which will later help in creating symmetric proofs for sets. Then we prove our main theorems by first proving the result for inner β -smoothings, and then showing that characterizations of the optimal general and outer smoothings follow from simple transformations.

Lemma 6 *Let K be a nonempty closed convex cone and C be a nonempty closed convex set. Then $K = C^\infty$ whenever $\text{dist}(C, K) < \infty$.*

⁴This fact can be verified by considering conjugates through the following equivalences:

$$f = g \iff f^* = g^* \iff f^* + \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 = g^* + \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \iff f \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 = g \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2.$$

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 [22, Corollary 9.1.2] $\square$

4.2.1 Proof of Theorem 4

Our proof is broken into several lemmas which combine together to immediately yield the theorem. Recall, due to Lemma 1, it suffices to consider only $\beta = 1$. We begin by showing in Lemma 7 that F_σ^{in} is an inner 1-smoothing that lower bounds any other inner 1-smoothing of σ .

Lemma 7 F_σ^{in} defined in (12) 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 (13) and the inclusion follows because $\text{epi } \varrho_\sigma = C_\sigma$ by Lemma 2. 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 (10) 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 (13) 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 2, 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 8 that f_σ^{in} is an inner 1-smoothing 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 8 f_σ^{in} defined in (12) 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

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

where the first equality follows by (13), 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

$$\text{dist}(f_\sigma^{\text{in}}, \sigma) = \sup_{x \in \mathcal{E}} f_\sigma^{\text{in}}(x) - \sigma(x)$$

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

where the second equality follows by Lemma 9 below, the inequality follows in part because of the subgradient inequality $\sigma(x) \geq \sigma(0) + \left\langle P_{\partial\sigma(0)}(x - x_\sigma), x \right\rangle$, and the last equality holds by definition of w_σ .

For the second part of the Lemma, first note that

$$\begin{aligned}
\text{dist}(\text{epi } f, \text{epi } \sigma) &= \sup_{x \in \mathcal{E}} \left\| (x, \sigma(x)) - P_{\text{epi } f}((x, \sigma(x))) \right\|_{\mathcal{E} \times \mathbb{R}} \\
&\leq \sup_{x \in \mathcal{E}} \left\| (x, \sigma(x)) - (x, f(x)) \right\|_{\mathcal{E} \times \mathbb{R}} \\
&= \sup_{x \in \mathcal{E}} |f(x) - \sigma(x)| \\
&= \text{dist}(f, \sigma)
\end{aligned}$$

where the first equality holds because (i) f is an inner smoothing, and (ii) only boundary points matter in computing the distance. As a result, 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 (13) and the last equality follows because Lemma 6 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. Note that

$$\begin{aligned}
\left(-\nabla f(0), f(0) - \frac{1}{2} \|\nabla f(0)\|_{\mathcal{E}}^2 \right) + \text{epi } \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 &= \text{epi} \left(f(0) - \frac{1}{2} \|\nabla f(0)\|_{\mathcal{E}}^2 + \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 + \nabla f(0) \right) \\
&= \text{epi} \left(f(0) + \langle \nabla f(0), \cdot \rangle + \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \right) \\
&\subseteq \text{epi } f \\
&\subseteq \text{epi } \sigma
\end{aligned} \tag{18}$$

where the first equality holds because $\text{epi}(r_0 + g(\cdot - x_0)) = (x_0, r_0) + \text{epi } g$ for any function $g : \mathcal{E} \rightarrow \mathbb{R}$ and any $(x_0, r_0) \in \mathcal{E} \times \mathbb{R}$, the second equality follows from completing the square, the first inclusion holds because $f(x) \leq f(0) + \langle \nabla f(0), x \rangle + \frac{1}{2} \|x\|_{\mathcal{E}}^2$ for all $x \in \mathcal{E}$ by (4), while the second follows by assumption. Our desired inclusion $(x_\sigma, r_\sigma) + \text{epi } \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2 \subseteq \text{epi } f$ then holds because of (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 $\left(-\nabla f(0), f(0) - \frac{1}{2} \|\nabla f(0)\|_{\mathcal{E}}^2 \right) \in C_\sigma$ by (18), while the second follows because $f(0) = |f(0) - \sigma(0)| \leq \text{dist}(f, \sigma) \leq w_\sigma$. Equation (10) 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 7 and 8, 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 7, and the second and third inequalities follow by assumption, while the last inequality by Lemma 8. Thus $\text{dist}(f, \sigma) = \text{dist}(F_\sigma^{\text{in}}, \sigma)$, hence f is optimal as Lemma 7 ensures F_σ^{in} is optimal. This concludes the proof of the second statement of the theorem. In addition, since F_σ^{in} is optimal and the above chain of inequalities proves $\text{dist}(F_\sigma^{\text{in}}, \sigma) = w_\sigma$, the third statement is also proved.

The first part of the theorem also follows directly from Lemmas 7 and 8. 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 8 and Lemma 1. 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 because w_σ is the optimal distance.

4.2.2 Proof of Theorems 3 and 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. We begin with smoothability claims. Since $f^{\text{in}} \geq \sigma$ and $\text{dist}(f^{\text{in}}, \sigma) \leq w_\sigma$ for any optimal inner 1-smoothing by Lemma 8, (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. Therefore, taking $r = \frac{w_\sigma}{2}$ we get that σ is $w_\sigma/2$ -smoothable. Similarly, taking $r = w_\sigma$ proves that σ is outer w_σ -smoothable. We now show that these parameters cannot be improved. For any 1-smoothing f , we have $f = f^{\text{in}} - \text{dist}(f, \sigma)$ where $f^{\text{in}} = f + \text{dist}(f, \sigma)$ 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 Theorem 4, 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 . This concludes the proof of smoothability claims for both Theorems 3 and 5. An immediate consequence of this is that $\text{dist}(f^{\text{gen}}, \sigma) = w_\sigma/2$ if f^{gen} is an optimal general 1-smoothing, and $\text{dist}(f^{\text{out}}, \sigma) = w_\sigma$ if f^{out} is an optimal outer 1-smoothing.

We now prove the “if and only if” characterization of optimal smoothings. This follows from Theorem 4, Equations (16) and the fact that (A) and (B) ensure that with $r = \frac{w_\sigma}{2}$ and $r = w_\sigma$, the transformation $f^{\text{in}} \mapsto f^{\text{in}} - r$ maps optimal inner 1-smoothings *on to* optimal general and outer 1-smoothings, respectively. Indeed, since the optimal general distance is $w_\sigma/2$ and $\max\{r, (w_\sigma - r)_+\} = r$ for $r = w_\sigma/2$, (A) shows that $f^{\text{in}} - r$ is an optimal general 1-smoothing if f^{in} is an optimal inner 1-smoothing, while (B) shows that for any optimal general 1-smoothing f^{gen} , we must have $f^{\text{gen}} = f^{\text{in}} - w_\sigma/2$ where f^{in} is an optimal inner 1-smoothing. An entirely analogous argument shows that with $r = w_\sigma$, the transformation $f^{\text{in}} \mapsto f^{\text{in}} - r$ maps optimal inner 1-smoothings *on to* optimal outer 1-smoothings. By combining this with Theorem 4 and Equations (16), the “if and only if” characterization follows.

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

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

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

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

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

4.3 Some Families of Optimal Smoothings

As an application of our theory, 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.

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

$$(\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2)(x) = \frac{1}{2} \|x\|_{\mathcal{E}}^2 - \frac{1}{2} \text{dist}(x, \partial\sigma(0))^2. \quad (19)$$

More generally, it follows that

$$((r_0 + \sigma(\cdot - x_0)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2)(x) = r_0 + \frac{1}{2} \|x - x_0\|_{\mathcal{E}}^2 - \frac{1}{2} \text{dist}(x - x_0, \partial\sigma(0))^2 \quad (20)$$

for all $x, x_0 \in \mathcal{E}$ and $r_0 \in \mathbb{R}$.

Proof First, we assert that

$$\operatorname{argmin}_{y \in \mathcal{E}} \sigma(y) + \frac{1}{2} \|x - y\|_{\mathcal{E}}^2 = x - P_{\partial\sigma(0)}(x). \quad (21)$$

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 (21). 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 $x - x^+ \in \partial\sigma(x^+)$. Hence x^+ satisfies the first-order optimality condition establishing (21). Then the claimed general formula for the Moreau envelope of σ follows as

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

where the second equality follows by (21), 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 $\operatorname{dist}(x, \partial\sigma(0)) = \|x - P_{\partial\sigma(0)}(x)\|_{\mathcal{E}}$, (19) follows. The general formula (20) follows immediately from (19) and the following calculation

$$\begin{aligned} ((r_0 + \sigma(\cdot - x_0)) \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2)(x) &= \inf_{y \in \mathcal{E}} r_0 + \sigma(y - x_0) + \frac{1}{2} \|x - x_0 + x_0 - y\|_{\mathcal{E}}^2 \\ &= r_0 + (\sigma \square \frac{1}{2} \|\cdot\|_{\mathcal{E}}^2)(x - x_0). \end{aligned}$$

□

4.3.1 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 [23–26]. 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 $\varrho_{\sigma} = r_{\sigma} + \sigma(\cdot - x_{\sigma})$, (17) ensures that the ReLU function has a unique optimal smoothing. In particular, from Theorem 3, calculating an infimal convolution of $\frac{1}{2} \|\cdot\|_{\mathcal{E}}^2$ with ϱ_{σ} gives the unique optimal 1-smoothing as

$$f_{\sigma}^{\text{gen}}(x) = \begin{cases} -\frac{1}{16} & \text{if } x < -1/2 \\ \frac{1}{2}(x + \frac{1}{2})^2 - \frac{1}{16} & \text{if } -1/2 \leq x \leq 1/2 \\ x - \frac{1}{16} & \text{otherwise.} \end{cases} \quad (22)$$

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

4.3.2 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 (16).

Proposition 10 *For any norm σ , it holds that $x_\sigma = 0$ and $w_\sigma = \varrho_\sigma(0)$. As a result, for any $\beta > 0$, 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 an inner smoothing are 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.*

Proof Since $x_\sigma = \text{argmin}_{x \in \mathcal{E}} \varrho_\sigma(x) + \frac{1}{2} \|x\|_\mathcal{E}^2$ by Equation (10), the first order condition $0 \in \partial \varrho_\sigma(0)$ is sufficient to prove $x_\sigma = 0$. Observe that (i) $\text{argmax}_{\zeta \in \partial \sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2 \subseteq \partial \varrho_\sigma(0)$ by Equation (9) 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 $\hat{\zeta} \in \text{argmax}_{\zeta \in \partial \sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2$, by convexity of the subdifferential. Since $\partial \sigma(0)$ is compact, it follows $\text{argmax}_{\zeta \in \partial \sigma(0)} \frac{1}{2} \|\zeta\|_\mathcal{E}^2$ is nonempty. Therefore $0 \in \partial \varrho_\sigma(0)$ and $x_\sigma = 0$ is proved. Since $w_\sigma = r_\sigma + \frac{1}{2} \|x_\sigma\|_\mathcal{E}^2$ by definition and $r_\sigma = \varrho_\sigma(x_\sigma)$ by Equations (10), it follows that $w_\sigma = r_\sigma = \varrho_\sigma(0)$ as claimed.

Now, note that

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

where the first equality holds by definition, the second holds because $(r + f) \square g = (f \square g) + r$ for any $r \in \mathbb{R}$ and two functions f and g , while the last equality holds because $x_\sigma = 0$ and $w_\sigma = r_\sigma$. Since $\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2 = \left[\sigma \square \frac{1}{2} \|\cdot\|_\mathcal{E}^2 \right]_{1/\beta}$ by positive homogeneity of σ , this proves optimality of $\sigma \square \frac{\beta}{2} \|\cdot\|_\mathcal{E}^2$, by Theorem 5. Optimal general and inner smoothings follow by Equations (16). $\square$

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 $\partial \sigma(0) = \text{conv}\{(\pm 1, \dots, \pm 1)\}$, 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 (17) ensures the one-norm has a unique optimal smoothing. Proposition 10 then gives the unique optimal outer 1-smoothing of

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

The error of approximation is $\text{dist}(f_\sigma^{\text{out}}, \sigma) = w_\sigma = \frac{d}{2}$. 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

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

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 (17). In this case, the largest (in function value) optimal outer 1-smoothing is given by

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

while the smallest optimal outer 1-smoothing is given by

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

One can check these piecewise formulae for $f \in \{f_\sigma^{\text{out}}, F_\sigma^{\text{out}}\}$ case-by-case over $x \in \mathbb{R}^2$ by verifying that $x - \nabla f(x)$ is a minimizer for the corresponding infimal convolution. For example, the four cases of f_σ^{out} correspond to the minimizer occurring at the origin, with only $|x_1|$ active, with only $2|x_2|$ active, or with both active. Therefore, a 1-smooth convex function f is an optimal 1-smoothing of this weighted infinity norm exactly if it lies between these two functions.

4.3.3 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 by⁵

$$f_\sigma^{\text{gen}}(x) = F_\sigma^{\text{gen}}(x) = \alpha + \frac{1}{2} \sum_{i=1}^J (x_{\pi(i)} - \alpha)^2 + \frac{1}{4} \left(1 + \frac{1}{d}\right) \quad (23)$$

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 A provides the resulting parallel formulas.

5 Optimal Smoothings of Convex Cones

We now characterize the set of all optimal smoothings of convex cones. Recall that $K \subseteq \mathcal{E}$ is a convex cone if

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

Assumption (Blanket Assumption about K) *For the rest of the paper, we assume K is a closed convex cone with a nonempty interior such that $K \neq \mathcal{E}$.*

Note that any closed convex cone K is β -smoothable for any $\beta > 0$ by, for example, considering the simple Minkowski sum $K + B(0, 1/\beta)$. As in the function setting, our

⁵This can be verified by directly computing $F_\sigma^{\text{gen}}(x) = \varrho_\sigma \square \frac{1}{2} \|\cdot\|_2^2 - w_\sigma/2 = (\frac{1}{2} + \sigma) \square \frac{1}{2} \|\cdot\|_2^2 - (1 - 1/d)/4$ using (21) and the formula for projecting on the unit simplex $\partial\sigma(0) = \text{conv}\{e_1, \dots, e_d\}$ given in [27, Figure 1]. Further note from this projection formula, this optimal smoothing can be computed in (expected) linear time in n [28].

focus here is not just demonstrating some smoothing S with $\text{dist}(S, K)$ finite, but rather in sets minimizing this gap.

Definition 7 A closed convex set S is an optimal β -smoothing of a closed convex cone K if S is β -smooth and $\text{dist}(K, S) \leq \text{dist}(K, \tilde{S})$ for all β -smooth, convex $\tilde{S}$.

As was the case for functions, Lemma 1 ensures that it suffices to characterize all optimal smoothings of K at the fixed smoothness level $\beta = 1$. To construct such smoothings, we use the following key objects, which we call the conic core, center, and width, respectively:

$$C_K := \{x \in \mathcal{E} \mid x + B(0, 1) \subseteq K\},$$

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

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

The conic core is closed and convex, as established by the following lemma which reformulates the conic core as an intersection of halfspaces.

Lemma 11 For any convex cone K , $C_K = \{x \in \mathcal{E} \mid \langle \zeta, x \rangle \leq -1, \forall \zeta \in N_K(0), \|\zeta\|_{\mathcal{E}} = 1\}$.

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 11, it is immediate that the conic center x_K , defined above as the projection of the origin onto C_K , exists and is unique. Lastly, we note that our conic width w_K is closely related to the conic width of Xiong and Freund [29, Section 1.3], which is an important quantity in the analysis of conic programming algorithms (see [29, 30]). In fact, it can easily be shown that their conic width is equal to $1/(w_K + 1) = 1/\|x_K\|_{\mathcal{E}}$.

5.1 Characterization of All Optimal Cone Smoothings

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

$$s_K^{\text{gen}} = \frac{1}{1+\frac{w_K}{2}} \left[(x_K + K) + B(0, 1 + \frac{w_K}{2}) \right], \quad S_K^{\text{gen}} = \frac{1}{1+\frac{w_K}{2}} \left[C_K + B(0, 1 + \frac{w_K}{2}) \right]. \quad (24)$$

The following theorem establishes that these are 1-smoothings and in fact provide the smallest and largest optimal 1-smoothings of K and suffice to characterize the set of all optimal smoothings.

Theorem 12 (Optimal Smoothings of K) *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 $[s_K^{\text{gen}}]_{1/\beta} \subseteq S \subseteq [S_K^{\text{gen}}]_{1/\beta}$. Further, any such S has $\text{dist}(S, K) = \frac{w_K}{2+w_K} \frac{1}{\beta}$.*

We defer the proof of this result to Section 5.2.2 as it follows as a consequence of Theorem 13.

5.1.1 Characterizations for Optimal Inner and Outer Cone Smoothings

Definition 8 A closed convex set S is an optimal inner β -smoothing of a closed convex cone K if S is β -smooth, $S \subseteq K$, and $\text{dist}(K, S) \leq \text{dist}(K, \tilde{S})$ for all β -smooth, convex $\tilde{S} \subseteq \mathcal{E}$ satisfying $\tilde{S} \subseteq K$. Analogously, a β -smooth, closed convex $S \supseteq K$ is an optimal outer β -smoothing of K if $\text{dist}(K, S) \leq \text{dist}(K, \tilde{S})$ for all β -smooth, closed convex $\tilde{S} \supseteq K$.

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

$$s_K^{\text{in}} = (x_K + K) + B(0, 1), \quad S_K^{\text{in}} = C_K + B(0, 1). \quad (25)$$

Theorem 13 formalizes this and is proven in Section 5.2.

Theorem 13 (Optimal Inner Smoothings of K) *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 $[s_K^{\text{in}}]_{1/\beta} \subseteq S \subseteq [S_K^{\text{in}}]_{1/\beta}$. Further, any such S has $\text{dist}(S, K) = \frac{w_K}{\beta}$.*

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

$$s_K^{\text{out}} = \frac{1}{1+w_K} [(x_K + K) + B(0, 1 + w_K)], \quad S_K^{\text{out}} = \frac{1}{1+w_K} [C_K + B(0, 1 + w_K)]. \quad (26)$$

Theorem 14 formalizes this and is proven in Section 5.2.2.

Theorem 14 (Optimal Outer Smoothings of K) *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 $[s_K^{\text{out}}]_{1/\beta} \subseteq S \subseteq [S_K^{\text{out}}]_{1/\beta}$. Further, any such S has $\text{dist}(S, K) = \frac{w_K}{1+w_K} \frac{1}{\beta}$.*

5.1.2 Symmetries between Function and Set Smoothings

First, we explicitly highlight the parallels between the optimal function and set smoothings. Noting that $\text{epi } \sigma$ is a convex cone, we have the following symmetry between the optimal inner smoothings of sublinear functions and convex cones.

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

The other optimal sets are generated from these inner smoothings. For functions, a simple translation gives the other optimal smoothings, as shown in (16). The parallel transformation for sets is a Minkowski addition together with a rescaling, giving

$$\begin{aligned} s_K^{\text{gen}} &= \frac{1}{1+\frac{w_K}{2}} [s_K^{\text{in}} + B(0, \frac{w_K}{2})], & S_K^{\text{gen}} &= \frac{1}{1+\frac{w_K}{2}} [S_K^{\text{in}} + B(0, \frac{w_K}{2})], \\ s_K^{\text{out}} &= \frac{1}{1+w_K} [s_K^{\text{in}} + B(0, w_K)], & S_K^{\text{out}} &= \frac{1}{1+w_K} [S_K^{\text{in}} + B(0, w_K)]. \end{aligned} \tag{27}$$

These reformulations are easy to verify from definitions.

Second, as was the case for functions, our theory provides an exact characterization of when a unique optimal smoothing exists.

$$K \text{ has unique smoothings} \iff s_K^{\text{in}} = S_K^{\text{in}} \iff x_K + K = C_K \tag{28}$$

where the first equivalence follows from our main results for sets and the second notes two closed convex sets C_1 and C_2 are equal if and only if $C_1 + B(0, 1) = C_2 + B(0, 1)$. This can be shown using support functions (see [19, Chapter 8E] for relevant concepts).

5.2 Proofs of Characterizations of Optimal Cone Smoothings

Our proofs follow the same pattern as the proofs in the function setting. We first prove the result for inner β -smoothings, and then show that characterizations of the optimal general and outer smoothings follow from simple transformations.

5.2.1 Proof of Theorem 13

The proof of this theorem follows the same style as that of Theorem 4. 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 15, proves the inclusion $S \subseteq S_K^{\text{in}}$ for any convex

inner 1-smoothing S . The second, Lemma 16, proves the inclusion $s_K^{\text{in}} \subseteq S$ for any optimal smoothing S .

Lemma 15 s_K^{in} defined in (25) 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

$$\begin{aligned} \text{dist}(s_K^{\text{in}}, K) &\geq \|P_{s_K^{\text{in}}}(0)\|_{\mathcal{E}} = \|P_{C_K}(0) + P_{B_1}(-P_{C_K}(0))\|_{\mathcal{E}} = \|x_K + P_{B_1}(-x_K)\|_{\mathcal{E}} \\ &= \|x_K\|_{\mathcal{E}} - 1 \end{aligned}$$

where the inequality holds because $0 \in K$, the first equality holds by Equation (1), the second equality holds by definition of x_K , while the last equality holds because $P_{B_1}(-x_K) = -\frac{x_K}{\|x_K\|_{\mathcal{E}}}$ as $K \neq \mathcal{E}$ implies $\|x_K\|_{\mathcal{E}} \geq 1$. Since $w_K = \|x_K\|_{\mathcal{E}} - 1$, the claimed $\text{dist}(s_K^{\text{in}}, K) \geq w_K$ holds. 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 16 s_K^{in} defined in (25) 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 16, we use the following useful intermediate lemma.

Lemma 17 $d(C; S + B(0, r)) = (d(C; S) - r)_+$ for any closed, convex sets C and S .

Proof of Lemma 17 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 (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 16 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 17, 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 6 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

$$\|x_K\|_{\mathcal{E}} \leq \|P_S(0) - \zeta\|_{\mathcal{E}} = \|P_S(0)\|_{\mathcal{E}} + 1 \leq \text{dist}(S, K) + 1 \leq w_K + 1 = \|x_K\|_{\mathcal{E}} \quad (29)$$

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 $-P_S(0) = 0 - P_S(0) = \|P_S(0)\|_{\mathcal{E}} \zeta$ since $0 - P_S(0) \in N_S(P_S(0))$ by convexity of S and $N_S(P_S(0)) = \{\alpha \zeta \in \mathcal{E} \mid \alpha \geq 0\}$ by smoothness of S ; the last inequality follows because $\text{dist}(S, K) \leq \text{dist}(s_K^{\text{in}}, K)$ by optimality assumption. Now, (29) 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 15 and 16, 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 15, 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 16. Thus $\text{dist}(S, K) = \text{dist}(S_K^{\text{in}}, K)$, hence S is optimal as Lemma 15 ensures S_K^{in} is optimal. This proves the second statement of the theorem. In addition, since S_K^{in} is optimal and the above chain of inequalities proves $\text{dist}(S_K^{\text{in}}, K) = w_K$, the third statement is also proved.

The first statement of Theorem 13 also follows immediately from Lemmas 15 and 16. 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 16 and 1. 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 15.

5.2.2 Proof of Theorems 12 and 14

The proof mirrors the proof of Theorems 3 and 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

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

(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 3 and 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, Equations (27), 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 1, $[S + B(0, r)]_{1/(1+r)}$ 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 17 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 1 then ensures $\text{dist} \left([S + B(0, r)]_{1/(1+r)}, K \right) \leq \frac{\max \{r, (\text{dist}(S, K) - r)_+\}}{1 + r}$.

We now prove (B). Suppose $\text{dist}(S, K) \leq \frac{r}{1+r}$. Then Lemma 1 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 17, while the last equality holds because K is invariant to positive rescalings.

5.3 Some Families of Optimal Cone Smoothings

By 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.

5.3.1 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

$$\begin{aligned} C_K &= \{x \in \mathbb{R}^d \mid \langle -e_i, x \rangle \leq -1, i = 1, \dots, d\} = e + K, \\ x_K &= \underset{x \in C_K}{\text{argmin}} \|x\|_2 = e, \\ w_K &= \|x_K\|_2 - 1 = \sqrt{d} - 1. \end{aligned}$$

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

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.

claimed quantities for the second-order and semidefinite cones where $N_K(0)$ is similarly generated as the conic hull of all $(u, -1)/\sqrt{2}$ or $-uu^T$ for unit vectors u , respectively.

We remark that more generally than these three examples, for any symmetric cone $K \subset \mathcal{E}$, the optimal smoothing is unique. We refer readers to [31] for a formal introduction to symmetric cones and here provide only an illustrative proof sketch: A symmetric cone K can be described as the cone of squares of a Euclidean Jordan Algebra $\mathcal{A}$ with identity e . Accordingly, consider the algebra's associated trace inner product and induced norm, and let $\lambda_{\min}(x)$ denote the smallest eigenvalue of some x . In this induced metric, the distance from any $x \in K$ to the boundary of K is its smallest eigenvalue. Substituting this fact into the definition of the conic core,

$$C_K = \{x \in \mathcal{E} \mid x + B(0, 1) \subseteq K\} = \{x \in \mathcal{E} \mid \lambda_{\min}(x) \geq 1\}.$$

Noting that $\lambda_{\min}(x - e) = \lambda_{\min}(x) - 1$, it follows that

$$x \in C_K \iff \lambda_{\min}(x - e) \geq 0 \iff x - e \in K.$$

Thus, $C_K = e + K$. Since $e \in \text{int}(K)$ and symmetric cones are self-dual, the projection of the origin onto the core is the identity $x_K = e$. Hence, one can conclude K has a unique optimal smoothing by (28).

5.3.2 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 [32].

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

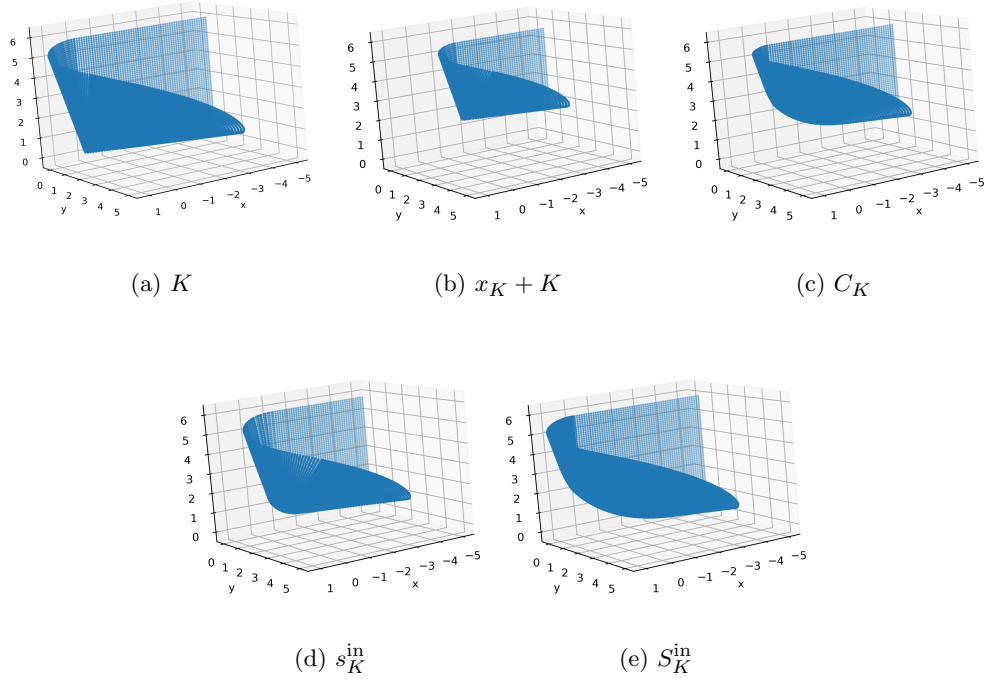


Fig. 2: The exponential cone, its translation $x_K + K$, its conic core, its minimal optimal 1-smoothing, and its maximal optimal 1-smoothing.

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 (28) 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 (25). 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

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

as Lemma 11 dictates. The supremum is numerically computed using a dense sample from the set $\{\zeta \in N_K(0) \mid \|\zeta\|_2 = 1\}$. 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.

6 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.

6.1 Beyond Sublinear Functions: Application to Amenable Functions

We consider the smoothing of convex functions g of the form $g = \sigma \circ G$, where σ is a sublinear function, under the assumption that

$$\|G(x) - G(y)\|_{\mathcal{E}'} \leq M \|x - y\|_{\mathcal{E}} \quad \text{and} \quad \|J(x) - J(y)\|_{\mathcal{E} \rightarrow \mathcal{E}'} \leq L \|x - y\|_{\mathcal{E}} \quad \forall x, y \in \mathcal{E}. \quad (30)$$

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 [33–35] and analysis [19, 36]. 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 [36, 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 (30) 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 18 *Consider any function $g = \sigma \circ G$ where σ is sublinear and M_σ -Lipschitz and $G : \mathcal{E} \rightarrow \mathcal{E}'$ satisfies (30). 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 convex 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

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

$$\leq (M_\sigma L + M^2 \beta') \|x - y\|_{\mathcal{E}}$$

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 $G, J(\cdot)^T$ 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 $\nabla f_\sigma(z) \in \partial\sigma(0)$ for all $z \in \mathcal{E}'$ by the following derivation.

$$(\nabla f_\sigma(z), -1) \in N_{\text{epi } f_\sigma}((z, f_\sigma(z))) \subseteq N_{(\text{epi } f_\sigma)^\infty}((0, 0)) = N_{\text{epi } \sigma}((0, 0))$$

where the ‘ $\subseteq$ ’ holds because $\nabla f_\sigma(z) \in \partial f_\sigma(z)$ by convexity of f_σ , the inclusion holds because we must have $\langle (\nabla f_\sigma(z), -1), (0, f_\sigma(0)) + (y, r) \rangle \leq \langle (\nabla f_\sigma(z), -1), (z, f_\sigma(z)) \rangle$ for all $(y, r) \in (\text{epi } f_\sigma)^\infty$ since $(0, f_\sigma(0)) + (y, r) \in \text{epi } f_\sigma$ by definition of $(\text{epi } f_\sigma)^\infty$, while equality holds because $\text{dist}(\sigma, f_\sigma) < \infty$ implies $\text{epi } \sigma = (\text{epi } f_\sigma)^\infty$ by Lemma 6.

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 $g = \sigma \circ G$. $\square$

Remark 1 By replacing f_σ^{gen} with f_σ^{in} , the same arguments can be used to show that $\sigma \circ G$ is inner $(M^2 w_\sigma, M_\sigma L)$ -smoothable. Similarly, f_σ^{out} can be used to show that $\sigma \circ G$ is outer $(M^2 w_\sigma, M_\sigma L)$ -smoothable.

6.1.1 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 -Lipschitz, Lemma 1 and Theorem 18 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 [37]), 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 inner 1-smoothing of σ . Rather the optimal inner smoothing is $f_\sigma^{\text{in}} = f_\sigma^{\text{gen}} + \frac{1}{4}(1 - \frac{1}{n})$ where f_σ^{gen} is given by (23). Using this smoothing and applying an accelerated gradient method to minimize $\varepsilon f_\sigma^{\text{in}}(G(x)/\varepsilon)$ 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 (23) and its gradient is somewhat more complex than those of the logSumExp, it can still be done in (expected) linear time in n [28].

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.

6.2 Beyond Convex Cones: Application to General Convex Sets

It is well-known that for any closed, 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 $C = \{x \in \mathcal{E} \mid (x - x_0, 1) \in K\}$; for example, one can take K to be the closure of the set $\{(x, r) \in \mathcal{E} \times \mathbb{R} \mid r > 0, x/r \in C - x_0\}$, for any $x_0 \in C$. Therefore, by smoothing K optimally, our theory can be used to approximate generic convex sets C . The following Theorem 19 shows that K can be chosen to be $(1 - \frac{R}{\sqrt{1+R^2}})$ -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 19 *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 $(1 - \frac{R}{\sqrt{1+R^2}})$ -smoothable.

Proof We only prove smoothability since the other statements are obvious. By Theorem 14, 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. $\square$

Appendix A 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 Frobenius 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

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

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 by⁶

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

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 $\max_{A \in \mathbb{S}^{d \times d}} |f_{\lambda_{\max}}^{\text{gen}}(A) - \sigma(A)| = \frac{1}{2} w_\sigma = \frac{1}{4}(1 - \frac{1}{d})$. 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 $\frac{1}{4}(1 - \frac{1}{d})/\beta$.

References

- [1] Moreau, J.-J.: Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France **93**, 273–299 (1965) <https://doi.org/10.24033/bsmf.1615>
- [2] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences **2**(1), 183–202 (2009) <https://doi.org/10.1137/080716542> <https://doi.org/10.1137/080716542>
- [3] Davis, D., Drusvyatskiy, D.: Stochastic model-based minimization of weakly convex functions. SIAM Journal on Optimization **29**(1), 207–239 (2019) <https://doi.org/10.1137/18M1178244> <https://doi.org/10.1137/18M1178244>
- [4] Penot, J.-P., Bougeard, M.L.: Approximation and decomposition properties of some classes of locally dc functions. Mathematical Programming **41**(1), 195–227 (1988)
- [5] Lasry, J.-M., Lions, P.-L.: A remark on regularization in hilbert spaces. Israel Journal of Mathematics **55**(3), 257–266 (1986)
- [6] Attouch, H., Aze, D.: 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, vol. 10, pp. 289–312 (1993). Elsevier
- [7] Ben-Tal, A., Teboulle, M.: A smoothing technique for nondifferentiable optimization problems. In: Dolecki, S. (ed.) Optimization, pp. 1–11. Springer, Berlin, Heidelberg (1989)

⁶This 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 (21) 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 .

- [8] Beck, A., Teboulle, M.: Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization* **22**, 557–580 (2012)
- [9] Sherman, J.: Nearly maximum flows in nearly linear time. In: *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. FOCS '13*, pp. 263–269. IEEE Computer Society, USA (2013). <https://doi.org/10.1109/FOCS.2013.36> . <https://doi.org/10.1109/FOCS.2013.36>
- [10] Nesterov, Y., Spokoiny, V.: Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics* **17**(2), 527–566 (2017) <https://doi.org/10.1007/s10208-015-9296-2>
- [11] Lei, M., Pong, T.K., Sun, S., Yue, M.-C.: Subdifferentially polynomially bounded functions and gaussian smoothing-based zeroth-order optimization. *arXiv preprint arXiv:2405.04150v3* (2024) [arXiv:2405.04150](https://arxiv.org/abs/2405.04150) [math.OC]
- [12] Diakonikolas, J., Guzmán, C.: Optimization on a finer scale: Bounded local subgradient variation perspective. *arXiv preprint arXiv:2303.05037v2* (2024) [arXiv:2403.16317](https://arxiv.org/abs/2303.05037) [math.OC]
- [13] John, F.: Extremum problems with inequalities as subsidiary conditions. In: *Studies and Essays, Presented to R. Courant on His 60th Birthday January 8, 1948*, pp. 187–204. Wiley Interscience, New York (1948)
- [14] Khachiyan, L.G.: Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research* **21**(2), 307–320 (1996)
- [15] Nesterov, Y.: Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimization Methods and Software* **23**(1), 109–142 (2008) <https://doi.org/10.1080/10556780701550059>
- [16] Luner, A., Grimmer, B.: Performance estimation for smooth and strongly convex sets. *arXiv preprint arXiv:2410.14811v2* (2024) [arXiv:2410.14811](https://arxiv.org/abs/2410.14811) [math.OC]
- [17] Deville, R., Fonf, V., Hájek, P.: Analytic and polyhedral approximation of convex bodies in separable polyhedral banach spaces. *Israel Journal of Mathematics* **105**(1), 139–154 (1998)
- [18] Xu, J., Pong, T.K., Sze, N.-s.: A smoothing moving balls approximation method for a class of conic-constrained difference-of-convex optimization problems. *arXiv preprint arXiv:2505.12314v1* (2025) [arXiv:2505.12314](https://arxiv.org/abs/2505.12314) [math.OC]
- [19] Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer, Berlin, Heidelberg (2009). <https://doi.org/10.1007/978-3-642-02431-3>
- [20] Liu, N., Grimmer, B.: Gauges and accelerated optimization over smooth and/or strongly convex sets. *arXiv preprint arXiv:2303.05037v3* (2024) [arXiv:2303.05037](https://arxiv.org/abs/2303.05037)

[math.OC]

- [21] Samakhoana, T., Grimmer, B.: Scalable projection-free optimization methods via multiradial duality theory. arXiv preprint arXiv:2403.13688v2 (2024) [arXiv:2403.13688](#) [math.OC]
- [22] Rockafellar, R.T.: Convex Analysis. Princeton landmarks in mathematics and physics. Princeton University Press, Princeton, N.J (1997 - 1970)
- [23] Barron, J.T.: Squareplus: A softplus-like algebraic rectifier. arXiv preprint arXiv:2112.11687v1 (2021) [arXiv:2112.11687](#) [cs.LG]
- [24] Ramachandran, P., Zoph, B., Le, Q.V.: Searching for activation functions. arXiv preprint arXiv:1710.05941v2 (2017) [arXiv:1710.05941](#) [cs.NE]
- [25] Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Gordon, G., Dunson, D., Dudík, M. (eds.) Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 15, pp. 315–323. PMLR, Fort Lauderdale, FL, USA (2011). <https://proceedings.mlr.press/v15/glorot11a.html>
- [26] Clevert, D.-A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289v5 (2016) [arXiv:1511.07289](#) [cs.LG]
- [27] Duchi, J., Shalev-Shwartz, S., Singer, Y., Chandra, T.: Efficient projections onto the l_1 -ball for learning in high dimensions. In: Proceedings of the 25th International Conference on Machine Learning, pp. 272–279 (2008)
- [28] Condat, L.: Fast projection onto the simplex and the l_1 ball. Math. Program. **158**(1–2), 575–585 (2016) <https://doi.org/10.1007/s10107-015-0946-6>
- [29] Xiong, Z., Freund, R.M.: The role of level-set geometry on the performance of pdhg for conic linear optimization. arXiv preprint arXiv:2406.01942v3 (2024) [arXiv:2406.01942](#) [math.OC]
- [30] Freund, R.M., Vera, J.R.: Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. SIAM Journal on Optimization **10**(1), 155–176 (1999) <https://doi.org/10.1137/S105262349732829X>
- [31] Faraut, J., Korányi, A.: Analysis on Symmetric Cones. Oxford Mathematical Monographs. Clarendon Press, Oxford (1994)
- [32] Boyd, S., Kim, S.-J., Vandenberghe, L., Hassibi, A.: A tutorial on geometric programming. Optimization and Engineering **8**(1), 67–127 (2007)

- [33] Burke, J.V.: Descent methods for composite nondifferentiable optimization problems. *Mathematical Programming* **33**(3), 260–279 (1985) <https://doi.org/10.1007/BF01584377>
- [34] Nesterov, Y.: Gradient methods for minimizing composite functions. *Mathematical Programming* **140**(1), 125–161 (2013) <https://doi.org/10.1007/s10107-012-0629-5>
- [35] Lewis, A.S., Wright, S.J.: A proximal method for composite minimization. *Mathematical Programming* **158**(1), 501–546 (2016) <https://doi.org/10.1007/s10107-015-0943-9>
- [36] Shapiro, A.: On a class of nonsmooth composite functions. *Mathematics of Operations Research* **28**(4), 677–692 (2003). Accessed 2025-04-01
- [37] Nesterov, Y.: A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Soviet Mathematics Doklady* **27**(2), 372–376 (1983)

Statements and Declarations

Funding Information. 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.

Competing Interests. There are no competing interests related to this research.

Author Contributions. All authors contributed equally to this work.

Data Availability. The manuscript has no associated data.