Journal of Nonlinear and Convex Analysis
Volume 9, Number 3, 2008, 379–406

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

CONVEX ANALYSIS BASED SMOOTH APPROXIMATIONS OF MAXIMUM FUNCTIONS AND SQUARED-DISTANCE FUNCTIONS

KEWEI ZHANG

Abstract. We use the explicit formula of the quadratic upper compensated convex transform of the maximum function $f_n(x_1, \dots, x_n) = \max_{1\le i\le n} x_i$ to study the properties of this $C^{1,1}$-smooth approximation of the maximum function, and apply a generalized quadratic lower compensated convex transform to the squared-distance function $\operatorname{dist}^2(x, K)$ to a finite set K. The maximum-like functions and squared-distance like functions are widely used in mathematical programming and applied mathematics. These transforms provide explicit convex $C^{1,1}$-smooth approximations with quadratic growth for these functions. We also use explicitly calculated examples of these approximations and their graphs to illustrate the effects of our compensated convex transforms.

1. Introduction

In this paper we first study analytic and geometric properties of the $C^{1,1}$ smooth approximations for maximum-like convex functions f by their quadratic upper compensated convex transform [45] (*quadratic upper transform* for short) defined by

$$C^u_{2,\lambda}(f(x)) = \lambda|x|^2 - C(\lambda|x|^2 - f(x)), \tag{1}$$

where $C(\lambda|x|^2 - f(x))$ is the convex envelope [31] of the function $\lambda|x|^2 - f(x)$. We compare our method with the prox-function regularization method proposed by Y. Nesterov [25]. Let

$$f_n(x) = \max_{1\le i\le n} x_i, \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n$$

be the n-dimensional maximum function. It was established in [45] that the quadratic upper transform $C^u_{2,\lambda}(f_n(x))$ can be calculated explicitly as

$$C^u_{2,\lambda}(f_n(x)) = \lambda|x|^2 - \lambda \operatorname{dist}^2\left(x,\, C\left(\frac{K_n}{2\lambda}\right)\right) + \frac{1}{4\lambda}, \quad x \in \mathbb{R}^n, \quad \lambda > 0, \tag{2}$$

where $\operatorname{dist}^2(x,\, C(K_n/(2\lambda))$ is the Euclidean squared distance function to the scaled canonical simplex $C(K_n/(2\lambda))$ defined by the convex hull of the finite set $K_n/(2\lambda) = \{e_1/(2\lambda), \dots, e_n/(2\lambda)\}$ with $\{e_1, \dots, e_n\}$ the standard Euclidean basis of $\mathbb{R}^n$. Let $\langle x, y\rangle$ be the standard Euclidean inner product of x, $y \in \mathbb{R}^n$ and $|x| = \sqrt{\langle x, x\rangle}$ the Euclidean norm of x. Sometimes we also use $x \cdot y$ to denote the same Euclidean inner product. We observe that the maximum function f_n can be written as $f_n(x) =$

2000 *Mathematics Subject Classification.* 90C25, 90C26, 49J52, 52A41, 65K10.
Key words and phrases. convex programming, maximum function, squared-distance function, quadratic upper and lower compensated convex transforms, $C^{1,1}$-smooth approximation, tight approximation.

$\max\{\langle x,y\rangle,\ y\in K_n\}$. If we let $J_n=(-K_n)\cup K_n$, then the l^∞ norm function $g_n(x)=\max_{1\le x_i\le n}|x_i|$ can be defined alternatively as $g_n(x)=\max\{\langle x,y\rangle,\ y\in J_n\}$. The common feature in these two functions is that the sets K_n and J_n are both contained in the unit sphere $S^{n-1}:=\{x\in\mathbb{R}^n,\ |x|=1\}$.

In [45], quadratic lower compensated convex transform was defined as

$$C^l_{2,\lambda}(f(x))=C(f(x)+\lambda|x|^2)-\lambda|x|^2, \tag{3}$$

It was established in [45] that the quadratic lower transform $C^l_{2,\lambda}(\mathrm{dist}^2(x,\,K))$ for the squared distance function $x\to\mathrm{dist}^2(x,\,K)$ to a compact set $K\subset\mathbb{R}^n$ gives rise to a $C^{1,1}$-'tight approximation' as $\lambda\to+\infty$. This type of lower translations were used before to squared-distance functions in the study of quasiconvex relaxations and gradient Young measure in the vectorial calculus of variations [40, 41, 42, 43, 44]. However, to derive explicit $C^{1,1}$-approximation formulae for squared-distance functions to finite sets, we need to modify the quadratic lower transform so that this can be achieved. We will discuss this after our study of the maximum-like functions and their quadratic upper transforms.

Theorem 1.1. *Let $F(x)=\max\{\langle x,y\rangle,\ y\in K\}$ where $K\subset S^{n-1}\subset\mathbb{R}^n$ is a given compact set. Then the quadratic upper transform $C^u_{2,\lambda}(F(x))$ defined by (1) is given by*

$$C^u_{2,\lambda}(F(x))=\lambda|x|^2-\lambda\,\mathrm{dist}^2\left(x,\,C\left(\frac{K}{2\lambda}\right)\right)+\frac{1}{4\lambda},\quad x\in\mathbb{R}^n,\quad \lambda>0. \tag{4}$$

Furthermore, we have

(i) **Uniform error estimate**

$$0\le C^u_{2,\lambda}(F(x))-F(x)\le\frac{1}{\lambda},\quad x\in\mathbb{R}^n,\quad\lambda>0. \tag{5}$$

(ii) **Smoothness and convexity**

$C^u_{2,\lambda}(F(x))$ is convex and belongs to $C^{1,1}(\mathbb{R}^n)$, and

$$|DC^u_{2,\lambda}(F(x))-DC^u_{2,\lambda}(F(y))|\le 2\lambda|x-y|,\quad |DC^u_{2,\lambda}(F(x))|\le 1,\quad x,\ y\in\mathbb{R}^n. \tag{6}$$

(iii) **Tightness of approximation**

Let M_K be the **medial axis** *of K (to be defined later), then*

$$C^u_{2,\lambda}(F(x))=F(x)\quad \textit{whenever}\quad x\in\mathbb{R}^n\setminus C(K/(2\lambda))\quad \textit{and}\quad \mathrm{dist}(x,\,M_K)>2/\lambda.$$

(iv) **Volume estimate for finite sets**

If we further assume that K is a finite set with $m>1$ elements, let $\bar B(0,R)\subset\mathbb{R}^n$ be the closed ball centred at 0 with radius $R>1/\lambda$ and $E^{(\lambda)}_R=\{x\in B(0,R),\ C^u_{2,\lambda}(F(x))\neq F(x)\}$. Then

$$\frac{\mathrm{meas}(E^{(\lambda)}_R)}{\mathrm{meas}(B(0,R))}\le\frac{C(m,n)}{\lambda R}$$

where $C(m,n)>0$ is a constant depending only on m and n.

The advantage of our approximation formula (4) is that it connects the original functions with the squared distance functions to parameterized (convex) sets involved, as seen in (4). Our approximation from above is tight in the sense that in any fixed ball centred at 0, the subset on which $F(x) \neq C^u_{2,\lambda}(F(x))$ is very small when $\lambda > 0$ is large. Intuitively, this means that the graph of $C^u_{2,\lambda}(F(x))$ is 'attached' to that of $F(x)$ from above, leaving the area where the values of $C^u_{2,\lambda}(F(x))$ and $F(x)$ are different very small, which are exactly a small neighbourhood of the set where $F(x)$ is not locally $C^{1,1}$. Note that in a general tightness result in [45] we showed that for any given continuous function $f : \mathbb{R}^n \mapsto \mathbb{R}$, satisfying the growth condition

$$|f(x)| \leq C_0|x|^2 + C_1, \quad x \in \mathbb{R}^n, \tag{7}$$

where C_0 and C_1 are non-negative constant, we have, at every point x where f is $C^{1,1}$ in a neighbourhood of x, that $C^u_{2,\lambda}(f(x)) = f(x)$ when $\lambda > \Lambda_x$ for some $\Lambda_x > 0$ large enough. Also $\lim_{\lambda\to\infty} C^u_{2,\lambda}(f(x)) = f(x)$ uniformly on any given compact set. In Theorem 1.1(iii), we know that the open set on which $F(\cdot)$ is locally $C^{1,1}$ is exactly $\mathbb{R}^n \setminus M_K$, thus we can obtain a more precise estimate of $\Lambda_x > 0$ for each $x \notin M_K$, such that $C^u_{2,\lambda}(F(x)) = F(x)$ when $\lambda > \Lambda_x$. If K is finite, Theorem 1.1(iv) provides a volume estimates for the set on which $C^u_{2,\lambda}(F(x)) \neq F(x)$ in a given ball $\bar{B}(0, R)$.

The maximum function f_n is one of the most important non-smooth convex functions in mathematical programming and its applications to economics and engineering [22, 27]. However, since the maximum function f_n is not smooth, many gradient based numerical methods cannot apply directly to such functions.

In the simple two-dimensional case, there are a number of smooth approximations for the maximum function [13, 28, 19]. For example, by using the fact that

$$f_2(x) = \max\{x_1, x_2\} = (|x_1 - x_2| + x_1 + x_2)/2,$$

one can easily think of the smooth approximation (see, for example [19])

$$F_2(\epsilon, x) = (\sqrt{(x_1 - x_2)^2 + \epsilon^2} + x_1 + x_2)/2.$$

for small $\epsilon > 0$. In the higher dimensional spaces, a well known C^∞-smooth approximation of the maximum function f_n called entropic regularization is given by

$$f_n(\lambda, x) = \frac{1}{\lambda}\log\left(\sum_{i=1}^n e^{\lambda x_i}\right), \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad \lambda > 0. \tag{8}$$

The so-called aggregation function $f_n(\lambda, x)$ converges to $f_n(x)$ uniformly as $\lambda \to +\infty$. In fact, there is a uniform error estimate depending on the dimension n as

$$0 \leq f_n(\lambda, x) - f_n(x) \leq \frac{1}{\lambda}\log n. \quad x \in \mathbb{R}^n \tag{9}$$

Note that in practice the above aggregation function is written in the form $\epsilon \log(\sum_{i=1}^n \exp(x_i/\epsilon))$, replacing λ by $1/\epsilon$ for small $\epsilon > 0$. The seemingly artificial formula (8) turns out to be a natural consequence of the prox-function regularization [25] by using the so-called entropic distance. To make the comparison between the entropic regularization and our quadratic upper transforms easier, we write the parameter as $\lambda > 0$ and the approximation converges to the original function when $\lambda \to +\infty$. In Example 2.10 and Example 2.11 at the end of Section 2, we plot graphs for these approximations. We compare the approximations of the maximum function by the entropic regularization method and by our upper transforms for one and two dimensional spaces.

The classical smoothing function (8) has been widely used in smooth minimizations for non-smooth convex programming problems [4, 9, 5, 21, 26, 29, 30, 14, 32, 34, 35, 25, 38]. The entropic regularization (8) also has a number of very useful algebraic and analytic properties which can be verified easily. We know that i) $f_n(\lambda, x)$ is invariant under the action of the permutation group S_n; ii) $f_n(\lambda, x)$ preserves the monotonicity property of the maximum function f_n; iii) the uniform error estimate (9) holds, and iv) the gradient satisfies $|Df_n(\lambda, x)| \leq 1$.

However, as pointed by several authors (see, for example [6, 25]) that for large $\lambda > 0$ (or equivalently, small $\epsilon = 1/\lambda$), calculations involving exponential functions in (8) can be very expensive and there is an issue concerning computational stability as the coupling of the exponential function followed by the logarithm function may involve dangerous operations. In [6], the authors reported that such a function may cause the so-called 'overflow problem'.

In order to overcome such a difficulty, a recursive extension of a simple two dimensional smoothing function was proposed in [6]. Let $\eta_\epsilon(x_1, x_2)$ be a smoothing approximation for the two dimensional maximum function, one may apply such a smoothing operation repeatedly. For example, in the three-dimensional case, the approximation $\eta_\epsilon(\eta_\epsilon(x_1, x_2), x_3)$ can be considered. This type of recursive extension preserves monotonicity property of the maximum function while the permutation invariance property of the original function is lost. Simple numerical examples by Matlab based programming presented in [6] show that numerical schemes based on (8) may blow up for not very large initial values while the recursive extension method still produces a correct answer. However, for large $n > 2$, it is difficult to write down an 'explicit' formula for this recursive extension. Also this approach requires several smoothing operations one after the other, the accumulated error of the approximation will depend on the dimension n.

It was established in [25, 3] that for $C^{1,1}$ convex functions, there are effective numerical methods for solving both unconstrained miniminzing problems in bounded convex domains of $\mathbb{R}^n$ and for constrained problems by using the Lagrangian multiplier method. Therefore an alternative simple geometric $C^{1,1}$ convex approximation such as (4) would be very useful for PC based programming.

Now we turn to the $C^{1,1}$ smooth approximation for the squared distance function $x \to \operatorname{dist}^2(x, K)$ for $x \in \mathbb{R}^n$ and a finite set K. The following squared-distance like function to a finite set was considered in [45]:

$$G(x) = \min_{1\leq i\leq m}(|x - y_i|^2 + b_i), \quad y_i \in \mathbb{R}^n, \quad b_i \geq 0, \ i = 1, 2, \dots, m.$$

Let $K_m = \{y_i, \ i = 1, 2, \dots, m\} \subset \mathbb{R}^n$, we see that if $b_i = 0$ for all $i = 1, 2, \dots,$, then $G(x)$ is exactly the squared distance function $\operatorname{dist}^2(x, K_m)$. The method used in [45] to find an explicit $C^{1,1}$ approximation for $G(x)$ is to consider the extended function $F_\lambda : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined by

$$F_\lambda(x, t) = \min_{1\leq i\leq m}\left((|x|^2 - 2x \cdot y_i + \frac{|y_i|^2}{1 + \lambda}) - 2t_i\right), \quad (x, t) \in \mathbb{R}^n \times \mathbb{R}^m.$$

If we define $\overline{|y|^2} = (|y_1|^2, \cdots, |y_m|^2)$ and $b = (b_1, \cdots, b_m) \in \mathbb{R}^m$, then it can be easily verified that $G(x) = F(x, -\lambda\overline{|y|^2}/(2(1+\lambda)) - b/2)$. If we define

$$g_\lambda(x,t) = \lambda(x,t) = \lambda|x|^2 + (1+\lambda)|t|^2), \quad (x,t) \in \mathbb{R}^n \times \mathbb{R}^m,$$

then the anisotropic quadratic lower transform $C^l_{g_\lambda}(F(x,t))$ is a $C^{1,1}$ approximation of $F_\lambda(x,t)$ which is given explicitly by

$$C_{g_\lambda}(F_\lambda(x,t)) = (1+\lambda)\,\mathrm{dist}^2\left((x,t), \frac{C(\hat{K}_m)}{1+\lambda}\right) - g_\lambda(x,t) - \frac{1}{1+\lambda},$$

where $\hat{K}_m = \{(y_i, e_i),\, i = 1, 2, \ldots, m\} \subset \mathbb{R}^n \times \mathbb{R}^m$ is a finite set. It was established in [45] that the restriction of $C_{g_\lambda}(F_\lambda(x,t))$ at $t = -\lambda\overline{|y|^2}|/(2(1+\lambda)) - b/2$ is a $C^{1,1}$-smooth approximation of $G(x)$. In fact if we let

$$G_\lambda(x) = C_{g_\lambda}\left(F_\lambda(x, -\lambda\overline{|y|^2}|/(2(1+\lambda)) - b/2)\right),$$

then we have

$$0 \le G(x) - G_\lambda(x) \le \frac{\mathrm{diam}^2(K_m)}{1+\lambda}, \quad |DG_\lambda(x)| \le 2|K_m|, \quad x \in \mathbb{R}^n,$$

where $\mathrm{diam}(K_m)$ and $|K_m| = \max\{|w|,\ w \in K_m\}$ are the diameter and norm of K_m respectively. This approximation can be considered as explicit if one can successfully calculates the squared distance function to the convex polytope $C(\hat{K}_m)/(1+\lambda)$.

In this paper we consider another example related to the squared-distance function defined by

$$H(x) = \sum_{j=1}^{m}(|x_j| - d_j)^2,\ x_j \in \mathbb{R}^n,\ x = (x_1, \cdots, x_m) \in \mathbb{R}^{mn},\ d_j > 0,\ j = 1, 2, \ldots, m.$$

The function $H(\cdot)$ arises from the so-called *molecular distance geometry problem* [24, 15, 18, 2] where the three-dimensional structure of a molecule is to be determined with the distances of a subset of the atoms involved are known. The problem is to find $x_1, \ldots, x_k \in \mathbb{R}^3$ which describe the positions of the atoms in the molecule such that for any given pair $(i,j) \in S$, one has $|x_i - x_j| = d_{ij}$, where S is a subset of all pairs (i,j). Obviously the required equalities holds if the function

$$f(x) = \sum_{(i,j)\in S} (|x_i - x_j| - d_{ij})^2, \quad x = (x_1, \ldots, x_k) \in \mathbb{R}^{3k}$$

reaches a minimizer x^* and $f(x^*) = 0$. An alternative model is $g(x) = \sum_{(i,j)\in S}(|x_i - x_j|^2 - d_{ij}^2)^2$. We only consider $H(\cdot)$ obtained from the first model function above.

Our plan for the rest of the paper is as follows. In Section 2, we establish Theorem 1.1 and other results related to quadratic upper transforms for maximum-like functions. We also use graphs of the absolute value function $f(x) = |x|$ and the maximum function $f(x,y) = \max\{x, y\}$ to compare approximations by our upper transforms and those by the entropic approximations given by (8). In Section 3, we derive a $C^{1,1}$-smooth approximation for the squared-distance-like function $H(x)$ defined above, followed by examples of some one and two dimensional explicitly calculated lower transforms for squared-distance-like functions. We also illustrate these examples of explicit approximations by comparing their graphs with those of the original functions by using Mathematica graphic softwares.

2. Explicit $C^{1,1}$-approximations for maximum-like function

We need some more notations and preliminaries. For a compact set $K \subset \mathbb{R}^n$ we define the diameter of the set K by $\mathrm{diam}(K) = \sup\{|x-y|,\ x,\, y \in K\}$ and let $l(K) = \sup\{x \in \mathbb{R}^n,\, x \in K\}$. The space $C^{1,1}(\mathbb{R}^n)$ consists of functions defined on $\mathbb{R}^n$ whose gradient Df are Lipschitz mappings:

$$|Df(x) - Df(y)| \le L_f|x-y|, \quad x,\ y \in \mathbb{R}^n$$

where $L_f \ge 0$ is the Lipschitz constant of Df.

Given a compact set $K \subset R^n$ and a fixed $t > 0$, let $tK = \{tx,\, x \in K\}$. For $\epsilon > 0$, we also let $(K)_\epsilon = \{x \in \mathbb{R}^n,\ \mathrm{dist}(x, K) \le \epsilon\}$ be the closed ϵ-neighbourhood of K. We denote by $C(K)$ its convex hull [31]. For a convex compact set $K \subset \mathbb{R}^n$ and a given point $x \in \mathbb{R}^n$, we denote by $P_K(x)$ the unique nearest point in K [20, 16] such that $\mathrm{dist}(x, K) = |x - P_K(x)|$. The mapping $x \to P_K(x)$ is called the convex projection to K. For a Lebesgue measurable set $\Omega \subset \mathbb{R}^n$, we denote by $\mathrm{meas}(\Omega)$ the Lebesgue measure of Ω.

Let S_n be the group of all permutations of the set $\{1,\dots,n\}$. If $\sigma \in S_n$ is given by $\sigma(1,\dots,n) = (i_1,\dots,i_n)$, then we define for a given point $x = (x_1,\dots,x_n)$, $\sigma(x) = (x_{i_1},\dots,x_{i_n})$ as permutation of component positions. It is easy to see that $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is both linear and isometric, satisfying that $|\sigma(x)| = |x|$. A function $g : \mathbb{R}^n \mapsto \mathbb{R}$ is called permutation invariant if $g(\sigma(x)) = g(x)$ for all $\sigma \in S_n$. For a vector $h = (h_1,\dots,h_n) \in \mathbb{R}^n$, we write $h \ge 0$ (respectively $h \le 0$) if $h_i \ge 0$ (respectively $h_i \le 0$) for all $i = 1,\dots,n$. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is monotone increasing (respectively, decreasing) if $f(x+h) \ge f(x)$ for all $x,\, h \in \mathbb{R}^n$ with $h \ge 0$ (respectively, if $f(x+h) \le f(x)$ for all $x,\, h \in \mathbb{R}^n$ with $h \le 0$). Observe that both the maximum function $f_n(x)$ defined above and the classical smooth approximation $f_n(\lambda, x)$ defined by (8) are permutation invariant and monotone increasing.

For a compact set $K \subset \mathbb{R}^n$, the *medial axis* $M_K \subset \mathbb{R}^n$ is the closed set on which $\mathrm{dist}^2(x, K)$ is not differentiable. It can be described alternatively as the closure of the set $M_K^0 \subset \mathbb{R}^n \setminus K$, where M_K^0 consists all of the points $x \in \mathbb{R}^n$ and a corresponding radius $r(x) > 0$ such that $B(x, r(x)) \subset \mathbb{R}^n \setminus K$ is the maximal open ball in the sense that any open ball B containing $B(x, r(x))$ cannot be completely contained in $\mathbb{R}^n \setminus K$ [7, 12, 37, 23]. A simple example of points $x \in M_K^0$ is that $\mathrm{dist}(x, K)$ is reached at least two different points on K. For a finite set $K \subset \mathbb{R}^n$, the medial axis of K is the so-called Voronoi diagram [36, 7, 12, 37]. In this case, M_K consists of those points whose distance to K can be reached by more than one points in K, that is,

$$M_K = \{x \in \mathbb{R}^n,\ \exists y_1,\, y_2 \in F,\ y_1 \ne y_2,\ \mathrm{dist}(x, K) = |x - x_1| = |x - y_2|\}.$$

Let us give a short description of the prox-function regularization method for smoothing a maximum like function proposed by Nesterov [25]. We will compare our method with this method later. Given a dual pair of finite dimensional normed linear spaces $(E,\, \|\cdot\|_1)$ and its dual $(E^*, \|\cdot\|_2)$. Let $A : E \to E^*$ be and linear map and $Q_2 \subset E^*$ be a compact convex set. Consider

$$f(x) = \max\{\langle Ax,\, y\rangle_2 - \phi(y),\ y \in Q_2\}$$

where $\langle\cdot,\,\cdot\rangle_2$ is a 'scalar product' on E^* and $\phi : Q_2 \to \mathbb{R}$ is a convex function.

A prox-function $d_2(y)$ is a strict convex function defined on $C(K)$ satisfying

$$\langle Dd_2(y_1) - Dd_2(y_2),\ y_1 - y_2\rangle \geq \sigma\|y_1 - y_2\|_2^2, \qquad \text{for some } \sigma > 0.$$

Let $y_0 = \arg\min_y\{d_2(y),\ y \in C(K)\}$ be the 'prox-centre', one may assume that $d_2(y_0) = 0$, hence $d_2(y) \geq \frac{\sigma}{2}\|y - y_0\|_2^2$. Let $\mu > 0$ be a small smoothness parameter, the smooth approximation $F_\mu(x)$ is defined by

$$f_\mu(x) = \max\{\langle Ax, y\rangle_2 - \mu d_2(y),\ y \in Q_2\}.$$

The resulting function is a $C^{1,1}$ approximation with Lipschitz constant of Df_μ given by $\|A\|^2/(\mu\sigma)$. In practice it was proposed in [25] to use either the standard Euclidean norm or the l^1 norm on E^*. We will compare this approach with ours later in the present paper.

We need the following two lemmas for the proof of Theorem 1.1.

Lemma 2.1. *Suppose $K \subset \mathbb{R}^n$ be a compact set, then*

(i) $C[\operatorname{dist}^2(x,\,K)] = \operatorname{dist}^2(x,\,C(K))$ *for all* $x \in \mathbb{R}^n$*;*

(ii) $0 \leq \operatorname{dist}^2(x,\,K) - \operatorname{dist}^2(x,\,C(K)) \leq \operatorname{diam}^2(K)$ *for all* $x \in \mathbb{R}^n$.

(iii) *If $K \subset \mathbb{R}^n$ is convex and compact, then*

$$\operatorname{dist}^2(x,\,K) = |x - P_K(x)|^2, \quad x \in \mathbb{R}^n, \tag{10}$$

where $P_K(x) \in K$ is the unique point in K such that (10) holds. Furthermore, $P_K : \mathbb{R}^n \to K$ is continuous and $D\operatorname{dist}^2(x,\,K) = 2(x - P_K(x))$, for all $x \in \mathbb{R}^n$.

Lemma 2.2. *Let $K \subset \mathbb{R}^n$ be a compact convex set and define*

$$f_\lambda(x) = \lambda|x|^2 - \lambda\operatorname{dist}^2\left(x,\,\frac{K}{2\lambda}\right) + C_\lambda$$

for $\lambda > 0$ with C_λ a constant. Let $P_{K/(2\lambda)}(x)$ the convex projection from $x \in \mathbb{R}^n$ to K. Then $f_\lambda(x) = 2\lambda\langle x,\,P_{K/(2\lambda)}(x)\rangle + C_\lambda$ and $f_\lambda(x)$ is both convex and of $C^{1,1}$ with

$$|Df_\lambda(x)| \leq l(K), \qquad |Df_\lambda(x) - Df_\lambda(y)| \leq 2\lambda|x - y|, \quad x,\ y \in \mathbb{R}^n.$$

Lemma 2.1 was proved in [45]. We give a proof to make the present paper self-contained.

Proof of Lemma 2.1. Item (i) is well-known. Even for the more general case of quasiconvex envelope of the p-distance function to a closed set $K \subset M^{N\times n}$ in the calculus of variations, we have (see [39]) that $Q[\operatorname{dist}^p(X,\,K)] = \operatorname{dist}^p(X,\,Q_p(K)]$ where $Q_p(K)$ is the quasiconvex hull if K is compact and p-quasiconvex hull of K is unbounded. The following is a short proof of (i).

Clearly, $\operatorname{dist}^2(x,K) \geq C\operatorname{dist}^2(x,K) \geq \operatorname{dist}^2(x,C(K))$. We prove the opposite inequality. Let $\operatorname{dist}^2(x,C(K)) = |x - x_0|^2$ for some $x_0 \in C(K)$. By Carathéodory's theorem [31], there are at most $n+1$ points $x_1,\ldots,x_{n+1} \in K$, such that $x_0 =$

$\sum_{i=1}^{n+1} \lambda_i x_i$ with $\lambda_i \geq 0$, $\sum_{i=1}^{n+1} \lambda_i = 1$. Thus by the convexity of $C\,\mathrm{dist}^2(\cdot, K)$, we have

$$\begin{aligned} C\,\mathrm{dist}^2(x, K) &= C\,\mathrm{dist}^2(x + \sum_{i=1}^{n+1} \lambda_i (x_i - x_0),\, K) \leq \sum_{i=1}^{n+1} \lambda_i C\,\mathrm{dist}^2(x + x_i - x_0,\, K) \\ &\leq \sum_{i=1}^{n+1} \lambda_i\,\mathrm{dist}^2(x + x_i - x_0,\, K) \leq \sum_{i=1}^{n+1} \lambda_i |x + x_i - x_0 - x_i|^2 \\ &= |x - x_0|^2 = \mathrm{dist}^2(x, C(K)). \end{aligned}$$

The proof of (i) is finished.

Item (ii) is a consequence of the Pythagorean theorem. Without loss of generality, we may assume that the affine dimension of $C(K)$ is n [31]. Clearly $\mathrm{dist}^2(x,\, C(K)) \leq \mathrm{dist}^2(x,\, K)$. Now given $x \in \mathbb{R}^n$ and $\mathrm{dist}^2(x,\, C(K)) < \mathrm{dist}^2(x,\, K)$ and $x \notin C(K)$, let $x_0 \in C(K) \setminus K$ be the unique point such that $\mathrm{dist}^2(x,\, C(K)) = |x - x_0|^2$. Note that for this particular x, $\mathrm{dist}^2(x,\, C(K)) = \mathrm{dist}^2(x,\, \partial C(K))$. Let $E \subset \mathbb{R}^n$ be a supporting plane of $C(K)$ passing through x_0 with the smallest dimension, then we see that $x - x_0$ is perpendicular to E. Now we take $x_1 \in E \cap K$ then we have

$$\begin{aligned} \mathrm{dist}^2(x,\, K) - \mathrm{dist}^2(x,\, C(K)) &\leq |x - x_1|^2 - |x - x_0|^2 \\ &= |x_1 - x_0|^2 \leq \mathrm{diam}^2(C(K)) = \mathrm{diam}^2(K). \end{aligned}$$

The proof for (ii) is finished.

Item (iii) is well known [23, 37]. The proof is also easy due to the nearest point property for compact convex sets [20]. So it is easy to see that $P_K(\cdot)$ is continuous. The formula for the gradient follows from the continuity of P_K and the definition of the squared distance function. We leave this last point to the proof of Lemma 2.2. □

Proof of Lemma 2.2. It is well-known that the convex projection $P_K : \mathbb{R}^n \to K$ to a closed convex set $K \subset \mathbb{R}^n$ satisfies (see [16])

$$|P_K(x) - P_K(y)|^2 \leq |x - y|^2 - |x - P_K(x)|^2 - |y - P_K(y)|^2, \quad x,\, y \in \mathbb{R}^n. \tag{11}$$

This implies that $|P_K(x) - P_K(y)| \leq |x - y|$ hence $P_K(\cdot)$ is a Lipschitz mapping with Lipschitz constant 1.

Now we consider $f_\lambda(\cdot)$. We have, by Lemma 2.1 that

$$Df_\lambda(x) = 2\lambda x - 2\lambda(x - P_{K/2\lambda}(x)) = 2\lambda P_{K/2\lambda}(x),$$

hence

$$|Df_\lambda(x)| = |2\lambda P_{K/2\lambda}(x)| \leq 2\lambda \max\left\{\frac{|y|}{2\lambda},\ y \in K\right\} = l(K). \tag{12}$$

By (11) we also have

$$|Df_\lambda(x) - Df_\lambda(y)| = 2\lambda |P_{K/2\lambda}(x) - P_{K/2\lambda}(y)| \leq 2\lambda |x - y|. \tag{13}$$

We are only left to show that f_λ is convex. Since $\lambda\,\mathrm{dist}^2(x,\,K/(2\lambda))$ is both convex and of $C^{1,1}$ we have, for any $x,\,y\in\mathbb{R}^n$ that

$$\begin{aligned}0 &\le \lambda\,\mathrm{dist}^2(x+y,\,K/(2\lambda)) - \lambda\,\mathrm{dist}^2(x,\,K/(2\lambda)) - \lambda D\,\mathrm{dist}^2(x,\,K/(2\lambda))\\ &= \lambda[|x+y-P_{K/(2\lambda)}(x+y)|^2 - |x-P_{K/(2\lambda)}(x)|^2 - 2\langle x-P_{K/(2\lambda)}(x),\,y\rangle]\\ &\le \lambda[|x+y-P_{K/(2\lambda)}(x)|^2 - |x-P_{K/(2\lambda)}(x)|^2 - 2\langle x-P_{K/(2\lambda)}(x),\,y\rangle] = \lambda|y|^2.\end{aligned}$$

Thus

$$\begin{aligned}&f_\lambda(x+y) - f_\lambda(x) - \langle Df_\lambda(x),\,y\rangle\\ &= \lambda|y|^2 - \lambda[\mathrm{dist}^2(x+y,\,K/(2\lambda)) - \mathrm{dist}^2(x,\,K/(2\lambda)) - D\,\mathrm{dist}^2(x,\,K/(2\lambda))] \ge 0,\end{aligned}$$

hence f_λ is convex. □

Note that if the convex set K consists of more than one point, (13) is a sharp estimate for all $\lambda > 0$ so that 2λ is the Lipschitz constant of $Df_\lambda(\cdot)$. This is clearly seen from (11) because for $x,\,y\in K/2\lambda$, $x\neq y$, we have $P_{K/2\lambda}(x) = x$, $P_{K/2\lambda}(y) = y$ and $Df_\lambda(x) = 2\lambda x$, $Df_\lambda(y) = 2\lambda y$.

Proof of Theorem 1.1. Let us establish (4) first. Note that $K\subset S^{n-1}$, we have

$$\begin{aligned}\lambda|x|^2 - F(x) &= \lambda|x|^2 - \max_{y\in K}\langle x, y\rangle\\ &= \min_{y\in K}[\lambda|x|^2 - \langle x, y\rangle] = \min_{y\in K}\left(\lambda\left|x-\frac{y}{2\lambda}\right|^2 - \frac{|y|^2}{4\lambda}\right)\\ &= \left(\lambda\min_{y\in K}\left|x-\frac{y}{2\lambda}\right|^2\right) - \frac{1}{4\lambda} = \lambda\,\mathrm{dist}^2\left(x,\,\frac{K}{2\lambda}\right) - \frac{1}{4\lambda}.\end{aligned}\tag{14}$$

Note that this identity also leads to

$$F(x) = \lambda|x|^2 - (\lambda|x|^2 - F(x)) = \lambda|x|^2 - \lambda\,\mathrm{dist}^2\left(x,\,\frac{K}{2\lambda}\right) + \frac{1}{4\lambda}.\tag{15}$$

Now by Lemma 2.1(i), we have

$$C[\lambda|x|^2 - F(x)] = \lambda\,\mathrm{dist}^2\left(x,\,C\left(\frac{K}{2\lambda}\right)\right) - \frac{1}{4\lambda},$$

so we have, by definition that

$$C^u_{2,\lambda}(F(x)) = \lambda|x|^2 - C[\lambda|x|^2 - F(x)] = \lambda|x|^2 - \lambda\,\mathrm{dist}^2\left(x,\,C\left(\frac{K}{2\lambda}\right)\right) + \frac{1}{4\lambda}.\tag{16}$$

Next we prove (i). Combining (15), (16) and applying Lemma 2.1(ii) we obtain

$$\begin{aligned}0 &\le C^u_{2,\lambda}(F(x)) - F(x)\\ &= \lambda\left[\mathrm{dist}^2\left(x,\,\frac{K}{2\lambda}\right) - \mathrm{dist}^2\left(x,\,C\left(\frac{K}{2\lambda}\right)\right)\right] \le \lambda\,\mathrm{diam}^2\left(\frac{K}{2\lambda}\right) \le \frac{1}{\lambda},\end{aligned}$$

as $\mathrm{diam}(K)\le 2$ in our case. The proof of (i) is finished.

Note that $l(K) = 1$ in our case. Thus (ii) follows from Lemma 2.2.

Now we prove (iii). Because $K\subset S^{n-1}$, we see that $M_K = M_{tK}$ for every $k > 0$. This is due to the fact that for any two point set $K_2 = \{x_1,\,x_2\}\subset\mathbb{S}^{n-1}$ with $x_1,\neq x_2$, we see that M_{K_2} is an $n-1$-dimensional subspace in $\mathbb{R}^n$ and M_{tK_2} is exactly the same subspace. Thus by a limit process we can easily see that in general $M_K = M_{tK}$ for $t > 0$ and $K \subset \mathbb{S}^{n-1}$.

For $x_0 \in \mathbb{R}^n \setminus C(K/(2\lambda))$ with $\operatorname{dist}(x_0,\, M_K) > 2/\lambda$, we only need to show that

$$C(\lambda|x_0|^2 - F(x_0)) = \lambda|x_0|^2 - F(x_0).$$

Note that (14) gives

$$\lambda|x|^2 - F(x) = \lambda \operatorname{dist}^2\left(x,\, \frac{K}{2\lambda}\right) - \frac{1}{4\lambda}$$

and $\lambda \operatorname{dist}^2(x,\, K/(2\lambda)) - 1/(4\lambda)$ is differentiable at x_0. In order to show that

$$C(\lambda|x_0|^2 - F(x_0)) = \lambda|x_0|^2 - F(x_0),$$

we only need to show that

$$[\lambda|x|^2 - F(x)] - [\lambda|x_0|^2 - F(x_0)] - \langle D(\lambda|x_0|^2 - F(x_0)),\, x - x_0\rangle \geq 0, \quad x \in \mathbb{R}^n,$$

or equivalently,

$$\text{(17)} \quad \operatorname{dist}^2\left(x,\, \frac{K}{2\lambda}\right) - \operatorname{dist}^2\left(x_0,\, \frac{K}{2\lambda}\right) - \left\langle D \operatorname{dist}^2\left(x_0,\, \frac{K}{2\lambda}\right),\, x - x_0\right\rangle \geq 0,\ x \in \mathbb{R}^n.$$

Consider the closed ball $\bar{B}(x_0, 2/\lambda) \subset \mathbb{R}^n \setminus M_K$, we see that for every $x \in B(x_0, 2/\lambda)$, there is a unique point $P(x) \in K/(2\lambda)$ such that $\operatorname{dist}(x,\, K/(2\lambda)) = |x - P(x)|$. Due to the uniqueness property of the nearest point $P(x)$ for $x \in \bar{B}(x_0, 2/\lambda)$, we can easily prove that $D \operatorname{dist}^2(x,\, K/(2\lambda)) = 2(x - P(x))$. Furthermore, we can prove that for $x \in \bar{B}(x_0, 2/\lambda)$, we have

$$|P(x) - P(y)| \leq |x - y|.$$

This last claim follows from the simple fact that for $x \in \bar{B}(x_0, 2/\lambda)$, $P(x) \in K/(2\lambda)$ is the only point such that $\langle z - P(x), x - P(x)\rangle \leq 0$ for all $z \in K/(2\lambda)$ [16]. Taking $z = P(y)$ for $y \in \bar{B}(x_0, 2/\lambda)$, we see that $|P(x) - P(y)|^2 \leq \langle P(x) - P(y),\, x - y\rangle$. The conclusion follows.

Now for $x \in \bar{B}(x_0, 2/\lambda)$, we have

$$\begin{aligned}
&\operatorname{dist}^2\left(x,\, \frac{K}{2\lambda}\right) - \operatorname{dist}^2\left(x_0,\, \frac{K}{2\lambda}\right) - \left\langle D \operatorname{dist}^2\left(x_0,\, \frac{K}{2\lambda}\right),\, x - x_0\right\rangle \\
&= \int_0^1 \left\langle D \operatorname{dist}^2\left(x_0 + t(x - x_0),\, \frac{K}{2\lambda}\right) - D \operatorname{dist}^2\left(x_0,\, \frac{K}{2\lambda}\right),\, x - x_0\right\rangle dt \\
&= 2\int_0^1 \langle (x_0 + t(x - x_0) - P(x_0 + t(x - x_0)) - (x_0 - P(x_0),\, x - x_0\rangle\, dt \\
&= 2\left[|x - x_0|^2 - \int_0^1 \langle P(x_0 + t(x - x_0) - P(x_0),\, x - x_0\rangle\, dt\right] \geq 0.
\end{aligned}$$

Thus (17) holds for $x \in \bar{B}(x_0, 2/\lambda)$. Next we consider $x \in \mathbb{R}^n \setminus B(x_0, 2/\lambda)$, we have $|x - x_0| \geq 2/\lambda$. We use the simple fact that $F(\cdot)$ is Lipschitz with Lipschitz constant

1, hence at x_0, $|DF(x_0)| \leq 1$. Then we have

$$\begin{aligned}&\operatorname{dist}^2\left(x, \frac{K}{2\lambda}\right) - \operatorname{dist}^2\left(x_0, \frac{K}{2\lambda}\right) - \left\langle D\operatorname{dist}^2\left(x_0, \frac{K}{2\lambda}\right), x - x_0\right\rangle \\ &= \lambda|x|^2 - F(x) - (\lambda|x_0|^2 - F(x_0)) - \langle 2\lambda x_0 - DF(x_0), x - x_0\rangle \\ &= \lambda|x - x_0|^2 - [F(x) - F(x_0) - \langle DF(x_0), x - x_0\rangle] \\ &= \lambda|x - x_0|^2 - 2|x - x_0| = |x - x_0|(\lambda|x - x_0| - 2) \geq 0,\end{aligned}$$

as $\lambda|x - x_0| \geq 2$. The proof of (iii) is finished.

Before we proceed, we remark that the medial axis M_K for a compact set $K \subset S^{n-1}$ is not necessarily a set of measure zero in $\mathbb{R}^n$. An example of $K \subset S^1$ is constructed in [23] such that the two-dimensional Lebesgue measure of M_K is nonzero, that is, $\operatorname{meas}(M_K) > 0$.

Next we prove (iv). Given any $x \in E_R$, that is, $|x| \leq R$, $x \notin C(K/(2\lambda))$ and $F(x) < C^u_{2,\lambda}(F(x))$, the last inequality is equivalent to $\operatorname{dist}^2(x, C(K/(2\lambda)) < \operatorname{dist}^2(x, K/(2\lambda))$.

Let $\Gamma^{(\lambda)}_k = \{\Delta^{(\lambda)}_{k,s}, s = 1, \dots s_k\}$ be the collection of k-dimensional faces (k-faces for short) of the polytope $C(K/(2\lambda)$ with $1 \leq k \leq n-1$. We define $\Delta^{(\lambda)}_{k,s}$ to be an open polytope in a k-dimensional space without boundary. We denote by s_k the number of such k-faces and let $l^{(m)}_k$ be the maximum number of k-faces a convex polytope with m-vertices (exposed points) [8]. The estimates of l_k and properties of convex polytopes can be found in [8]. Now we see that

$$E^{(\lambda)}_R = \cup_{k=1}^{n-1} \cup_{s=1}^{s_k} V^{(\lambda,R)}_{k,s},$$
$$\text{where} \quad V^{(\lambda,R)}_{k,s} = \{x \in E_R, \operatorname{dist}(x, C(K/(2\lambda))) = \operatorname{dist}(x, \Delta^{(\lambda)}_{k,s}).$$

This is easy to see as $\operatorname{dist}(x, C(K/(2\lambda))) = |x - P_{C(K/(2\lambda))}(x)|$. Now we can easily give an easy rough estimate of $\operatorname{meas}(V^{(\lambda,R)}_{k,s})$. The k-dimensional face $\Delta^{(\lambda)}_{k,s}$ is contained in a k-dimensional ball with radius at most $\operatorname{diam}(K/(2\lambda)) \leq 1/\lambda$, so $\operatorname{meas}_k(\Delta^{(\lambda)}_{k,s}) \leq b_k/\lambda^k$ where b_k is the volume of the k-dimensional unit ball. Thus $\operatorname{meas}(V^{(\lambda,R)}_{k,s}) \leq b_k b_{n-k} R^{n-k}/\lambda^k$. Therefore

$$\operatorname{meas}(E^{(\lambda)}_R) \leq \sum_{k=1}^{n-1}\sum_{s=1}^{s_k} \operatorname{meas}(V^{(\lambda,R)}_{k,s}) \leq \sum_{k=1}^{n-1} s_k b_k b_{n-k} \frac{R^{n-k}}{\lambda^k},$$

which implies that

$$\begin{aligned}\frac{\operatorname{meas}(E^{(\lambda)}_R)}{\operatorname{meas}(B_k)} &\leq \sum_{k=1}^{n-1} s_k b_k b_{n-k} \frac{R^{n-k}}{b_n R^n \lambda^k} \\ &\leq \sum_{k=1}^{n-1} s_k b_k b_{n-k} \frac{1}{b_n (R\lambda)^k} \leq \frac{1}{R\lambda}\sum_{k=1}^{n-1} l^{(m)}_k \frac{b_k b_{n-k}}{b_n} = \frac{C(m,n)}{R\lambda}. \qquad \square\end{aligned}$$

Our next result is concerned with the general set $K \subset \mathbb{R}^n$ and we have to modify our function to obtain a simple approximation formula.

Theorem 2.3. *Let* $G(x) = \max\{\langle x, y\rangle,\ y \in K\}$ *where* $K \subset \mathbb{R}^n$ *is a non-empty compact set. Let*

$$G_\lambda(x) = \max\{\langle x, y\rangle - \frac{|y|^2}{4\lambda},\ y \in K\}, \quad \lambda > 0. \tag{18}$$

Then

$$G(x) - \frac{l^2(K)}{4\lambda} \leq G_\lambda(x) \leq G(x), \quad x \in \mathbb{R}^n \tag{19}$$

and the quadratic upper transform of $G_\lambda(\cdot)$ *is given by*

$$C^u_{2,\lambda}(G_\lambda(x)) = \lambda|x|^2 - \lambda \operatorname{dist}^2\left(x,\, C\left(\frac{K}{2\lambda}\right)\right), \quad x \in \mathbb{R}^n, \quad \lambda > 0. \tag{20}$$

Furthermore,

$$-\frac{l^2(K)}{4\lambda} \leq C^u_{2,\lambda}(G_\lambda(x)) - G(x) \leq \frac{\operatorname{diam}^2(K)}{4\lambda}, \quad x \in \mathbb{R}^n, \tag{21}$$

$C^u_{2,\lambda}(G_\lambda(x))$ *is convex and belongs to* $C^{1,1}(\mathbb{R}^n)$*, and*

$$|DC^u_{2,\lambda}(G_\lambda(x)) - DC^u_{2,\lambda}(G_\lambda(y))| \leq 2\lambda|x-y|, \qquad |DC^u_{2,\lambda}(G_\lambda(x))| \leq l(K), \quad x,\ y \in \mathbb{R}^n. \tag{22}$$

Proof of Theorem 2.3. The error estimate (19) is easy to obtain. Now we prove (20). We have, for $\lambda > 0$ that

$$\begin{aligned}\lambda|x|^2 - G_\lambda(x) &= \lambda|x|^2 - \max_{y\in K}[\langle x,\, y\rangle - \frac{|y|^2}{4\lambda}] \\ &= \min_{y\in K}[\lambda|x|^2 - \langle x,\, y\rangle + \frac{|y|^2}{4\lambda}] = \lambda \operatorname{dist}^2\left(x,\, \frac{K}{2\lambda}\right).\end{aligned}$$

Thus we can rewrite $G_\lambda(x)$ as

$$G_\lambda(x) = \lambda|x|^2 - (\lambda|x|^2 - G_\lambda(x)) = \lambda|x|^2 - \lambda \operatorname{dist}^2\left(x,\, \frac{K}{2\lambda}\right). \tag{23}$$

Now we have

$$C(\lambda|x|^2 - G_\lambda(x)) = \lambda \operatorname{dist}^2\left(x,\, C\left(\frac{K}{2\lambda}\right)\right),$$

hence

$$C^u_{2,\lambda}(G_\lambda(x))) = \lambda|x|^2 - \lambda \operatorname{dist}^2\left(x, C\left(\frac{K}{2\lambda}\right)\right). \tag{24}$$

Applying Lemma 2.1(ii) to (23) and (24) we obtain

$$0 \leq C^u_{2,\lambda}(G_\lambda(x))) - G_\lambda(x) \leq \frac{\operatorname{diam}^2(K)}{4\lambda}. \tag{25}$$

Combining (19) and (25) we obtain the error estimate (21). Claim (22) of Theorem 2.3 follow from Lemma 2.2. □

Let us compare the prox-function regularization method for smoothing a maximum like function proposed by Nesterov [25] with the general quadratic upper transform method stated in Theorem 2.3. Let us consider the maximum like function $G(x)$ in Theorem 2.3 as an example. Applying to the maximum function $f_n(x) = \max_{1\le i\le n} x_i = \langle x, y\rangle,\ y \in K_n\} = \max\{\langle x, y\rangle,\ y \in C(K_n)\}$, explicit calculations are performed for $d_2(y) = \log n + \sum_{i=1}^n y_i \log y_i$ in $C(K_n)$ which is a simplex. The calculations recovers the entropy smoothing formula $F_\mu(x) = \mu \log(\sum_{i=1}^n e^{y_i/\mu})$. The case under the standard Euclidean norm with $d_2 = \frac{1}{2}\sum_{i=1}^n (y_i - 1/n)^2$ was not calculated in [25] for the maximum function and left as an abstract smoothing formula. This last prox-function was used once for smoothing the Euclidean norm in [25].

The advantage of the prox-function method is that it can be defined for more general functions in the form $f(x) = \max\{\langle x, y\rangle - \phi(y),\ y \in C(K)\}$ where ϕ is a convex function. If the set $C(K)$ is simple, the norm of E^* can be chosen so that $d_2(\cdot)$ can be other than the squared Euclidean distance function to the prox-centre. However, for a given convex set $C(K)$, the prox-centre itself might not be easy to find and to define a proper prox-function requires some insight of the set $C(K)$. On the other hand, Theorem 2.3 reduces the smooth approximation to the calculation of the squared distance function to a shrinking family of known convex sets. Also in the definition of quadratic upper transform, one only needs to consider the squared Euclidean norm instead of searching a prox-function $d_2(\cdot)$ based on the geometry of $C(K)$. In fact it was established in [45] that for a general convex function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying the growth condition (7), the quadratic upper transform $C^u_{2,\lambda}(f(x))$ is both $C^{1,1}$ and convex with Lipschitz constant of the gradient at most 8λ. Clearly, the prox-function method does not apply to general convex functions. Now in the case of Theorem 2.3, if we rewrite $C^u_{2,\lambda}(G_\lambda(x))$, we can find a close connection between the quadratic upper transform of $G_\lambda(x)$ and the prox-function approximation of $G(\cdot)$. We have, by Theorem 2.3 that

$$C^u_{2,\lambda}(G_\lambda(x)) = \lambda|x|^2 - \lambda\,\mathrm{dist}^2\left(x,\, C\left(\frac{K}{2\lambda}\right)\right) = \lambda|x|^2 - \lambda \min_{y\in C(K)}\left[\left|x - \frac{y}{2\lambda}\right|^2\right]$$
$$= \lambda \max_{y\in C(K)}\left[|x|^2 - \left|x - \frac{y}{2\lambda}\right|^2\right] = \lambda \max_{y\in C(K)}\left[\langle x,\, \frac{y}{\lambda}\rangle - \frac{|y|^2}{4\lambda^2}\right] = \max_{y\in C(K)}\left[\langle x,\, y\rangle - \frac{|y|^2}{4\lambda}\right].$$

Thus the quadratic upper transform in this case is ‘almost’ the prox-function smoothing function with $d_2(y) = |y|^2/2$ and $\mu = 1/(2\lambda)$ by using the standard Euclidean norm. However, it is not obvious from the prox-function smoothing formula to obtain $C^u_{2,\lambda}(G_\lambda(x))$ through a backward calculation.

As the first application of Theorem 1.1, we have the following result concerning our $C^{1,1}$ convex approximation to the maximum function $f_n(x) = \max_{1\le i\le n} x_i$. Since f_n is both monotone increasing and permutation invariant, we can derive some more detailed properties of $C^u_{2,\lambda}(f_n(x))$. As before, let $e_1, \dots, e_n$ be the standard Euclidean basis of $\mathbb{R}^n$ and let $K_n = \{e_1, \dots, e_n\}$. Note here that geometrically K_n consists all the vertices of an $n-1$-dimensional simplex in $\mathbb{R}^n$.

Corollary 2.4. *Let $f_n(x) = \max_{1\leq i\leq n} x_i$ be the maximum function, then the quadratic upper transform is given by*

$$C^u_{2,\lambda}(f_n(x)) = \lambda|x|^2 - \lambda\,\mathrm{dist}^2\left(x,\, C\left(\frac{K_n}{2\lambda}\right)\right) + \frac{1}{4\lambda}, \quad x \in \mathbb{R}^n, \quad \lambda > 0. \tag{26}$$

Items (i)–(iv) in Theorem 1.1 hold for f_n and K_n with a better error estimate than that in Theorem 1.1 as

$$0 \leq C^u_{2,\lambda}(f_n(x)) - f_n(x) \leq \frac{1}{2\lambda}. \tag{27}$$

Furthermore,

(i) *The quadratic upper transform $C^u_{2,\lambda}(f_n(\cdot))$ given by (26) is both monotone increasing and permutation invariant.*
(ii) *Given $x \in M_{K_n}$ where M_{K_n} is the Voronoi diagram of K_n such that $\mathrm{dist}(x, K_n) = |x - e_{m_j}|$ for $j = 1, \dots, s \leq n$, where $\{e_{m_1}, \dots, e_{m_s}\}$ is a subset of K_n consisting of at least two points, then*

$$\lim_{\lambda\to+\infty} DC^u_{2,\lambda}(f_n(x)) = \frac{1}{s}\sum_{j=1}^{s} e_{m_j} \in \partial f_n(x),$$

where $\partial f_n(x)$ is the subdifferential of $f_n(x)$ [11].

Questions concerning the limit of gradient of the smooth approximations to the subdifferential of the original function such as Item (ii) in Corollary 2.4 was studied in [16]. We need the following Lemma to prove Corollary 2.4(i).

Lemma 2.5. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $|f(x)| \leq C_0|x|^2 + C_1$ and is monotone increasing (respectively, decreasing), then $C^u_{2,\lambda}(f(x))$ is monotone increasing (respectively, decreasing).*

Proof. Suppose f is monotone increasing and fix $h \in \mathbb{R}^n$, $h \geq 0$. Let $g_h(x) = f(x+h)$, we have $-f(x) \geq -g_h(x)$ for all $x \in \mathbb{R}^n$, hence for any fixed $\lambda > C_0$,

$$\begin{aligned} C[\lambda|x|^2 - f(x)] &\geq C_x[\lambda|x|^2 - g_h(x)] = C_x[\lambda|x+h|^2 - f(x+h) - 2\lambda x\cdot h - \lambda|h|^2] \\ &= C_x[\lambda|x+h|^2 - f(x+h)] - 2\lambda x\cdot h - \lambda|h|^2, \end{aligned}$$

due to the fact that $l_h(x) := -2\lambda x\cdot h - \lambda|h|^2$ is affine in x, where $C_x(F(x,h))$ is the convex envelope in x for functions F depending on both x and h. It is easy to see that

$$C_x[\lambda|x+h|^2 - f(x+h)] = C[\lambda|x+h|^2 - f(x+h)]$$

where the convex envelope on the right hand side of the above is taken with respect to the $x+h$-variable. Thus

$$C[\lambda|x|^2 - f(x)) \geq C[\lambda|x+h|^2 - f(x+h)] - 2x\cdot h - |h|^2,$$

which implies that

$$\begin{aligned} C^u_{2,\lambda}(f(x)) &\leq \lambda|x|^2 - C[\lambda|x+h|^2 - f(x+h)] + 2\lambda x\cdot h + 2\lambda|h|^2 \\ &= \lambda|x+h|^2 - C[\lambda|x+h|^2 - f(x+h)] = C^u_{2,\lambda}[f(x+h)]. \end{aligned}$$

The proof for monotone decreasing f is similar. □

Proof of Corollary 2.4. Estimate (27) follows from the fact that $\operatorname{diam}(K_n) = \sqrt{2}$. We only need to prove Items (i) and (ii). All other claims are consequences of Theorem 1.1 with $K = K_n \subset \mathbb{S}^{n-1}$. Now we show that $C^u_{2,\lambda}(f_n(\cdot))$ is permutation invariant, that is, for every permutation $\sigma \in S_n$, $C^u_{2,\lambda}(f_n(\sigma(x))) = C^u_{2,\lambda}(f_n(x))$ for all $x \in \mathbb{R}^n$, where $\sigma(x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Since $\lambda|x|^2$ is clearly permutation invariant, we only need to show that

$$\operatorname{dist}^2\left(\sigma(x),\, C\left(\frac{K}{2\lambda}\right)\right) = \operatorname{dist}^2\left(\sigma(x), \frac{K}{2\lambda}\right).$$

This is also easy to see. Note that $\operatorname{dist}^2(x, C(K/2)) = |x - P_{C(K/(2,\lambda))}(x)|^2$ and the convex projection is a convex combination of the vertices of the simplex $C(K_n/(2\lambda))$. Thus we can write $P_{C(K/(2,\lambda))}(x) = \sum_{i=1}^n \tau_i e_i$ with $\tau_i \geq 0$ and $\sum_{i=1}^n \tau_i = 1$. Let $\tau = (\tau_1, \ldots, \tau_n)$, then

$$\begin{aligned}
&\operatorname{dist}^2\left(x, C\left(\frac{K}{2\lambda}\right)\right)\\
&= |x - P_{C(K/(2\lambda))}(x)|^2 = |x - \tau|^2 = |\sigma(x - \tau)|^2 = |\sigma(x) - \sigma(\tau)|^2\\
&= |\sigma(x) - P_{C(K/(2\lambda))}(\sigma(x))|^2 = \operatorname{dist}^2(\sigma(x), C(K/(2\lambda)),
\end{aligned}$$

as $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is both linear and isometric under the Euclidean norm.

Since f_n is monotone increasing, we have, by Lemma 2.5 that $C^u_{2,\lambda}(f_n(\sigma(x)))$ is monotone increasing.

Now we prove Item (ii). Given $x \in M_{K_n}$ and let $e_{i_1}, \ldots, e_{i_k} \in K_n$ $(2 \leq k \leq n)$ be such that $\operatorname{dist}(x,\, K_n) = |x - e_{i_1}| = \cdots = |x - e_{i_k}|$. Then it is easy to see that

$$\operatorname{dist}\left(x,\, C\left(\frac{K_n}{2\lambda}\right)\right) = \left|x - \frac{e_{i_1}}{2\lambda}\right| = \cdots = \left|x - \frac{e_{i_k}}{2\lambda}\right|$$

and

$$P_{C(K_n/(2\lambda))}(x) = \frac{1}{k}\sum_{s=1}^k \frac{e_{i_s}}{2\lambda}, \quad DC^u_{2,\lambda}(f_n(x)) = 2\lambda P_{C(K_n/(2\lambda))}(x) = \frac{1}{n}\sum_{s=1}^k e_{i_s}$$

for all $\lambda > 0$ hence

$$\lim_{\lambda\to\infty} DC^u_{2,\lambda}(f_n(x)) = \frac{1}{k}\sum_{s=1}^k e_{i_k} \in \partial f_n(x) = C(\{e_{i_1}, \ldots, e_{i_k}\}).$$

The last equality can be found, for example, in [25]. □

Since the explicit formula (26) for $C^u_{2,\lambda}(f_n(\cdot))$ is not in a closed form, we would like to remark on the explicit calculations of $C^u_{2,\lambda}(f_n(x))$ and $DC^u_{2,\lambda}(f_n(x))$. These all come down to the calculation of $P_{C(K_n/(2\lambda))}(x)$. However, we have

$$P_{C(K_n/(2\lambda))}(x) = \frac{1}{2\lambda} P_{C(K_n)}(2\lambda x)$$

due to the fact that

$$\begin{aligned}\operatorname{dist}\left(x,\, C\left(\frac{K_n}{2\lambda}\right)\right) &= \inf\left\{\left|x-\frac{y}{2\lambda}\right|,\ y\in C(K_n)\right\}\\ &= \frac{1}{2\lambda}\inf\{|2\lambda x - y|,\ y\in C(K_n)\}\\ &= \frac{1}{2\lambda}|2\lambda x - P_{C(K_n)}(2\lambda x)| = \left|x-\frac{1}{2\lambda}P_{C(K_n)}(2\lambda x)\right|,\end{aligned}$$

hence we only need to find an effective way of calculating $P_{C(K_n)}(x)$. Now we follow [33]. After an re-ordering of the components by a permutation σ_x such that $\sigma_x(x) = (x_{i_1},\dots,x_{i_n})$ with $x_{i_1}\geq x_{i_2}\geq\cdots\geq x_{i_n}$. Then the k-th component of $P_C(K_n)(\sigma_x(x))$ is given by

$$[P_{C(K_n)}(\sigma_x(x))]_k = \begin{cases} x_{i_k}+\frac{1}{m^*}\left(1-\sum_{s=1}^{m^*}x_{i_s}\right), & 1\leq k\leq m^*,\\ 0, & k>m^*,\end{cases}$$

where $m^*\leq n$ is the largest positive integer m such that $S^{(m)} := \sum_{s=1}^{m}(a_{i_s}-a_{i_m})\leq 1$. It was proved in [33] that the sequence $S^{(m)}$ is non-decreasing and $S^{(1)} = 0$. Thus

$$P_{C(K_n)}(x) = \sigma_x^{-1}(P_{C(K_n)}(\sigma_x(x))).$$

Note also that $DC^u_{2,\lambda}(f_n(x)) = P_{C(K_n/(2\lambda))}(x)$.

Next we consider $C^{1,1}$ smooth approximations to maximum-like functions. The first example of such a function is the l_∞ norm on $\mathbb{R}^n$ defined by

$$g_n(x) = \|x\|_\infty = \max_{1\leq i\leq n}|x_i|,\quad x=(x_1,x_2,\dots,x_n)\in\mathbb{R}^n. \tag{28}$$

It is known that if we enlarge the dimension of the space, we can write $g_n(x) = \{x_1,-x_1,\dots,x_n,-x_n\}$ and define the C^∞ approximation (see, for example [25]) by

$$g_n(\lambda,x) = \frac{1}{\lambda}\log\left(\sum_{i=1}^{n}\left[e^{\lambda x_i}+e^{-\lambda x_i}\right]\right).$$

However, if we apply the quadratic upper transform to $g_n(\cdot)$ we do not have to enlarge the dimension. Instead, we consider a cube with twice as many points as the set K_n by defining $J_n = K_n\cup(-K_n)$, where K_n is defined in Corollary 2.4. Observe that J_n is the set of vertices of an n-dimensional cube with side-length $\sqrt{2}$. Again, we can derive a very simple geometric formula for $C^u_{2,\lambda}(g_n(x))$ without enlarge the dimension of the space.

Corollary 2.6. *Suppose $g_n(\cdot)$ is defined by (28) and J_n is defined as above, then*

$$C^u_{2,\lambda}(g_n(x)) = \lambda|x|^2-\lambda\operatorname{dist}^2\left(x,\, C\left(\frac{J_n}{2\lambda}\right)\right)+\frac{1}{4\lambda}, x\in\mathbb{R}^n,\quad \lambda>0. \tag{29}$$

Items (i)–(iv) of Theorem 1.1 hold. In particular, if we replace the ball $B(0,R)$ in Theorem 1.1(iv) by the cube $D_R = RC(J_n) = \{Ry,\ y\in C(J_n)\}$ and let

$$U_R^{(\lambda)} = \{x\in C(RJ_n)\setminus C(J_n/(2\lambda)),\ g_n(x)<C^u_{2,\lambda}(g_n(x))\},$$

then we have

$$\frac{\operatorname{meas}(U_R^{(\lambda)})}{\operatorname{meas}(D_R)} = 1 - \left(1 - \frac{1}{2\lambda R}\right)^n. \tag{30}$$

Furthermore, for any $x \in M_{J_n}$ *where* M_{J_n} *is the Voronoi diagram of* J_n *such that* $\operatorname{dist}(x, J_n) = |x - u_{m_j}|$ *for* $j = 1, \dots, s \leq 2n$*, where* $\{u_{m_1}, \dots, u_{m_s}\}$ *is a subset of* J_n *consisting of at least two points, then*

$$\lim_{\lambda \to +\infty} DC^u_{2,\lambda}(g_n(x)) = \frac{1}{s}\sum_{j=1}^{s} u_{m_j} \in \partial g_n(x).$$

Proof of Corollary 2.6. We only establish the volume ratio equality (30) for $\operatorname{meas}(U_R^{(\lambda)}))/\operatorname{meas}(D_R)$. The set $D_R \setminus U_R^{(\lambda)}$ on which $C^u_{2,\lambda}(g_n(x)) = g_n(x)$ consists of 2^n smaller cubes with total volume $2^n(\sqrt{2}R - \sqrt{2}/(2\lambda))^n$. Thus

$$\operatorname{meas}(U_R^{(\lambda)})) = \left((\sqrt{2}R)^n - 2^n(\frac{\sqrt{2}R}{2} - \frac{\sqrt{2}}{2(2\lambda)}\right)^n,$$

hence

$$\frac{\operatorname{meas}(U_R^{(\lambda)}))}{\operatorname{meas}(D_R)} = \frac{(\sqrt{2}R)^n - (\sqrt{2}R - \sqrt{2}/(2\lambda))^n}{(\sqrt{2}R)^n} = 1 - \left(1 - \left(\frac{1}{2R\lambda}\right)^n\right). \qquad \square$$

In our next example we consider the positive part of the maximum function defined by $f_+(x) = \max\{f_n(x), 0\}$, where f_n is the maximum function in Corollary 2.4. If we apply the quadratic upper transform to f_+, we will see that the calculation can be more complicated than that for the maximum function. However, if we give away a bit of tightness of the approximation, we may perturb our original function f_+ by a family of functions $f_\lambda(x) = \max\{f(x), 1/(4\lambda)\}$. It is easy to see that $0 \leq f_\lambda(x) - f_+(x) \leq 1/(4\lambda)$. Thus for large $\lambda > 0$, $f_\lambda(\cdot)$ is a very good uniform approximation of $f_+(\cdot)$. Now we denote by $K_n^{(0)} = K_n \cup \{0\}$, where K_n is defined in Corollary 2.4. The proof of the following result follows directly from that of Theorem 1.1 and is left to interested readers.

Corollary 2.7. *Let* f_λ *and* $K_n^{(0)}$ *be defined above, then for* $\lambda > 0$*,*

$$C^u_{2,\lambda}(f_\lambda(x)) = \lambda|x|^2 - \lambda \operatorname{dist}^2\left(x, C\left(\frac{K_n^{(0)}}{2\lambda}\right)\right) + \frac{1}{4\lambda}, x \in \mathbb{R}^n. \tag{31}$$

Theorem 1.1(i)–(iv) are satisfied for f_λ *and* $C^u_{2,\lambda}(f_\lambda(x))$*. Furthermore, we have the following uniform error estimate*

$$0 \leq C^u_{2,\lambda}(f_\lambda(x)) - f_+(x) \leq \frac{3}{4\lambda}, \quad x \in \mathbb{R}^n. \tag{32}$$

Our next example is a tight $C^{1,1}$ approximation for the non-smooth function

$$G(y,t) = \max\{|y_i| + t_i,\ y = (y_1, \dots, y_m) \in \mathbb{R}^N,\ y_i \in \mathbb{R}^{k_i},\ t = (t_1, \dots, t_m) \in \mathbb{R}^m\}, \tag{33}$$

where $N = \sum_{i=1}^m k_i$. The function $G(y,t)$ arises from the so-call second order cone optimization where the following constrained minimization problem is considered [1, 10]

$$\begin{cases} \text{minimize} \quad f\cdot x, \\ \text{subject to} \quad |A_i x + b_i| \le c_i \cdot x + d_i, \quad i = 1,\dots,m, \end{cases} \tag{34}$$

where $x \in \mathbb{R}^n$ is independent variable, and all other quantities, $f \in \mathbb{R}^n$, $A_i \in M^{k_i\times n}$, $b_i \in \mathbb{R}^{k_i}$, $c_i \in \mathbb{R}^n$, and $d_i \in \mathbb{R}$ are given. As the standard Euclidean norm is not smooth at 0, the constraints are non-smooth. We can rewrite the constraint in an equivalent form

$$\max_{1\le i\le m} [|A_i x + b_i| - (c_i \cdot x + d_i)] \le 0, \quad x \in \mathbb{R}^n. \tag{35}$$

Now we consider the function $G(x,t)$ defined above by (33) on $\mathbb{R}^{N+m} = \Pi_{i=1}^m(\mathbb{R}^{k_i} \times \mathbb{R})$ Then (35) is just the evaluation of $G(y,t)$ at $y_i = A_i x + b_i$, $t_i = -(c_i \cdot x + d_i)$. Before we write down the upper compensated convex transforms $C^u_{2,\lambda}(G(y,t))$, let us define some subsets of $\mathbb{R}^{N+m} = \Pi_{i=1}^m \mathbb{R}^{k_i} \times \mathbb{R}$. We define the sphere

$$S^i_\lambda = \{(y,t) \in \mathbb{R}^{N+m}, \quad y_j = 0,\ t_j = 0,\ j \ne i, \quad |y_i| = \frac{1}{2\lambda},\ t_i = \frac{1}{2\lambda}\}, \quad S_\lambda = \cup S^i_\lambda.$$

The following is an explicit $C^{1,1}$ approximation of $G(y,t)$ by its quadratic upper transform.

Theorem 2.8. *The quadratic upper transform $C^u_{2,\lambda}(G(y,t))$ is given by*

$$C^u_{2,\lambda}(G(y,t)) = \lambda(|y|^2 + |t|^2) - \lambda\,\mathrm{dist}^2((y,t),\ C(S_\lambda)) + \frac{1}{2\lambda}, \quad (y,t) \in \mathbb{R}^{N+m}, \quad \lambda > 0. \tag{36}$$

Furthermore, $C^u_{2,\lambda}(G(y,t))$ is convex and belongs to $C^{1,1}(\mathbb{R}^{N+m})$,

$$0 \le C^u_{2,\lambda}(G(y,t)) - G(y,t) \le \frac{1}{\lambda}, \quad (y,t) \in \mathbb{R}^{N+m}, \quad \lambda > 0, \tag{37}$$

and

$$|DC^u_{2,\lambda}(G(y,t))| \le 1,\ |DC^u_{2,\lambda}(G(y,t)) - DC^u_{2,\lambda}(G(y',t'))| \le 2\lambda|(y,t) - (y',t')|. \tag{38}$$

The proof of Theorem 2.8 is similar to that of Theorem 1.1. The calculation of $C^u_{2,\lambda}(G(y,t))$ is straight forward as

$$\begin{aligned} \lambda(|y|^2 + |t|^2) - G(y,t) &= \min_{1\le i\le m} [\lambda(|y|^2 + |t|^2) - (|y_i| + t_i) + \frac{1}{2\lambda}] - \frac{1}{2\lambda} \\ &= \lambda\,\mathrm{dist}^2((y,t),\ S_\lambda) + \frac{1}{2\lambda}. \end{aligned}$$

From the above we see that the calculation of $C^u_{2,\lambda}(G(y,t))$. However, if one tries to apply the prox-function regularization, it is not obvious how to find a proper $d_2(\cdot)$ and calculate a corresponding smooth approximation $G_\mu(y,t)$.

Next we apply Theorem 2.3 to the following piecewise affine convex function widely used in non-smooth convex programming [25, 17]:

$$h(x) = \min_{1\le j\le m}(a_j \cdot x + b_j), \quad x,\ a_j \in \mathbb{R}^n,\ b_j \in \mathbb{R}. \tag{39}$$

The sub-level sets of such a function can also be used to define convex polytopes [17]. A convex polytope K is a convex body in $\mathbb{R}^n$ generated by a finite set. Therefore K has finitely many facets, that is, faces of dimension $n-1$. Let $V_1, \dots, V_m$ be the half spaces defining K with $V_i = \{x,\ a_i \cdot x + b_i \le 0\}$, where $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, 2, \dots m$. We can write $K = \cap_{i=1}^m V_i$ [31]. Thus K can also be written as

$$K = \{x \in \mathbb{R}^n,\ \max_{i=1}^{m}(a_i \cdot x + b_i) \le 0\}.$$

Thus if we can find an explicit smooth approximation of $h(x) := \max_{i=1}^m (a_i \cdot x + b_i)$, then we can see that we may obtain a $C^{1,1}$-approximation of the convex polytope.

One way to smooth h is to apply Theorem 1.1 directly to the m-dimensional maximum function $f_m(y)$ and calculate $C^l_{2,\lambda}(f_m(y))$, then we set $y_i = a_i \cdot x + b_i$. Alternatively, we can introduce another perturbation of h and calculate the resulting upper transform. Let $b = (b_1, \dots, b_m)$ and $c_\lambda = (c_1^{(\lambda)}, \cdots, c_1^{(\lambda)}) \in \mathbb{R}^m$ with $c_i^{(\lambda)} = |a_j|^2/(4\lambda)$, $i = 1, 2, \dots$. Let

$$H_\lambda(x,t) = \max_{1\le j\le m}\left(a_j \cdot x - \frac{|a_j|^2}{4\lambda} + t_j\right), \quad (x,t) \in \mathbb{R}^n \times \mathbb{R}^m, \quad t = (t_1, \cdots, t_m).$$

We have $h(x) \ge H_\lambda(x,\ b + c_\lambda)$. Let $L_m = \{a_j,\ 1 \le j \le m\}$ and $\hat{L_m} = \{(a_j, e_j),\ 1 \le j \le m\}$ where $\{e_1, \dots, e_m\}$ is the standard basis of $\mathbb{R}^m$. Then we have

Theorem 2.9. *Under the assumptions above, we have for $\lambda > 0$ that*

$$\begin{aligned}
&C^u_{2,\lambda}(H_\lambda(x,t)) = \lambda(|x|^2 + |t|^2) - \lambda\,\mathrm{dist}^2\left((x,t), C\left(\frac{L_m}{2\lambda}\right)\right),\\
&C^u_{2,\lambda}(H_\lambda(x,b+c_\lambda)) = \lambda(|x|^2 + |b+c_\lambda|^2) - \lambda\,\mathrm{dist}^2\left((x,b+c_\lambda), C\left(\frac{\hat{L_m}}{2\lambda}\right)\right),\\
&h(x) \le C^u_{2,\lambda}(H_\lambda(x,b+c_\lambda)) \le h(x) + \frac{\mathrm{diam}^2(\hat{L_m})}{2\lambda},\\
&|DC^u_{2,\lambda}(H_\lambda(x,b+c_\lambda))| \le |L_m|, \quad x \in \mathbb{R}^n.
\end{aligned} \tag{40}$$

We conclude this section by showing the graphs of two explicit examples of maximum-like functions and their approximations.

Example 2.10. Consider the absolute value function $f(x) = |x|$, we see that $f(x) = \max\{-x,\ x\}$. Thus we have the entropic regularization for $f(x)$ given by the aggregation function (8), that is,

$$f(\lambda, x) = \frac{1}{\lambda}\log\left(e^{\lambda x} + e^{-\lambda x}\right), \quad x \in \mathbb{R}, \quad \lambda > 0.$$

We also have

$$C^u_{2,\lambda}(f(x)) = \begin{cases} \lambda x^2 + \dfrac{1}{4\lambda}, & |x| \le \dfrac{1}{2\lambda},\\[2ex] f(x), & |x| \ge \dfrac{1}{2\lambda}.\end{cases}$$

Clearly, $C^u_{2,\lambda}(f(x))$ is a tight approximation of $f(x)$ while $f(\lambda, x)$ is not.

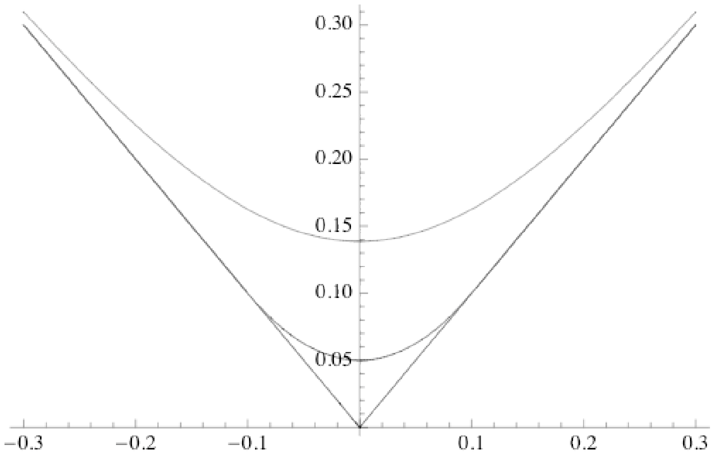

FIGURE 1. (Example 2.10) $f(x) = |x|$ (bottom), $C^u_{2,\lambda}(f(x))$ (middle) and $f(\lambda, x)$ (top) with $\lambda = 5$

In Figure 1 we plot the graphs for $f(x) = |x|$ (bottom), $C^u_{2,\lambda}(f(x))$ (middle) and $f(\lambda, x)$ (top) for $\lambda = 5$ to compare different approximations to $f(x)$ from above by $C^u_{2,\lambda}(f(x))$ and $f(\lambda, x)$.

Example 2.11. Now we illustrate approximations to the maximum function $f(x, y) = \max\{x,\ y\}$ in $\mathbb{R}^2$ by both our upper transform $C^u_{2,\lambda}(f(x, y))$ and by the entropic method.

We may write $C^u_{2,\lambda}(f(x, y))$ in the following explicit form

$$C^u_{2,\lambda}(f(x,y)) = \begin{cases} f(x,y), & |x-y| \geq \dfrac{1}{2\lambda}, \\ \dfrac{\lambda}{2}(x-y)^2 + \dfrac{1}{2}(x+y) + \dfrac{1}{8\lambda}, & |x-y| \leq \dfrac{1}{2\lambda}. \end{cases}$$

The entropic approximation (8) in this case is $f(\lambda, x, y) = (1/\lambda)\log[e^{\lambda x} + e^{\lambda y}]$.

In Figure 2, we display graphs for $f(x, y)$, $C^u_{2,\lambda}(f(x, y))$ and $f(\lambda, x, y)$ in the domain $|x| \leq 0.2$, $|y| \leq 0.2$ with $\lambda = 4$. In Figure 3, we examine the errors of approximations with $\lambda = 4$ in the order from the left to the right: $C^u_{2,\lambda}(f(x, y)) - f(x, y)$, $f(\lambda, x, y) - f(x, y)$ and $f(\lambda, x, y) - C^u_{2,\lambda}(f(x, y))$. We can see that the maximum error between $C^u_{2,\lambda}(f(x, y))$ and $f(x, y)$ in the domain $|x| \leq 0.2$, $|y| \leq 0.2$ is about 0.03 while the gap between $f(\lambda, x, y)$ and $f(x, y)$ is about 0.2. The maximum different between $f(\lambda, x, y)$ and $C^u_{2,\lambda}(f(x, y))$ is close to 0.2. Therefore we see that $C^u_{2,\lambda}(f(x, y))$ is a better approximation than $f(\lambda, x, y)$ for the same λ.

3. $C^{1,1}$ APPROXIMATIONS FOR SQUARED-DISTANCE AND DISTANCE LIKE FUNCTIONS

In this section we first define a smooth approximation for the function $H(x) = \sum_{j=1}^m (|x_j| - d_j)^2$. Then we examine some one and two dimensional examples for explicitly calculated lower transforms for squared distance like functions. We use

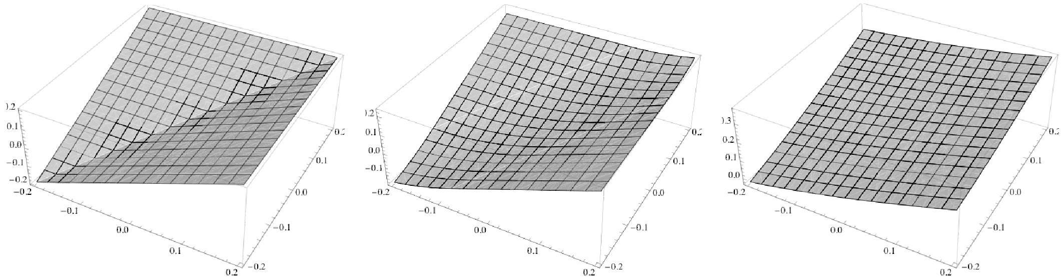

FIGURE 2. (Example 2.11) Left: $f(x,y)$, middle: $C^u_{2,\lambda}(f(x,y))$ and right: $f(\lambda,x,y)$ with $\lambda=4$.

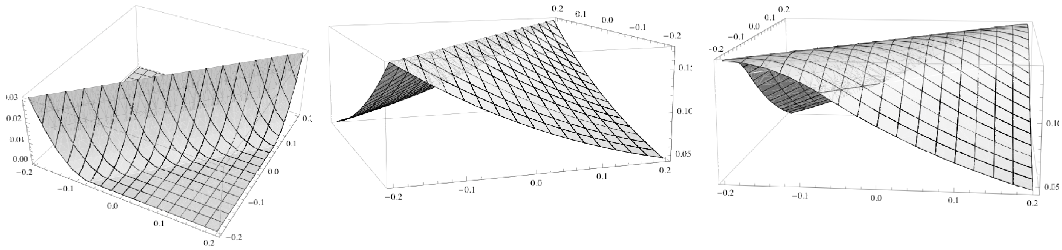

FIGURE 3. (Example 2.11) Left: $C^u_{2,\lambda}(f(x,y))-f(x,y)$, middle: $f(\lambda,x,y)-f(x,y)$ and right: $f(\lambda,x,y)-C^u_{2,\lambda}(f(x,y))$ with $\lambda=4$.

Mathematica again to plot graphs of these approximations which illustrate the effects of the so called tight approximations. We also consider an approximation to the distance function by a combination of lower and upper transforms.

It is easy to see that $(|x_j|-d_j)^2=\mathrm{dist}^2(x_j,\,S_{d_j})$ where $S_{d_j}=\{y\in\mathbb{R}^n,\ |y|=d_j\}$ is the $n-1$-dimensional sphere centred at 0 with radius d_j. Let $T\subset\mathbb{R}^{nm}$ be the torus $T=\{y=(y_1,\dots,y_m),\ y_i\in S_{d_i}\}$, we see that $H(x)=\sum_{j=1}^m(|x_j|-d_j)^2=\mathrm{dist}^2(x,\,T)$. Now we consider

$$\lambda|x|^2+H(x)=\sum_{j=1}^m[\lambda|x_j|^2+(|x_j|-d_j)^2]=\sum_{j=1}^m\left[(1+\lambda)\left(|x_j|-\frac{d_j}{1+\lambda}\right)^2+\frac{\lambda d_j^2}{1+\lambda}\right].$$

This indicates that if we consider another function

$$H_\lambda(x,t)=\sum_{j=1}^m(|x_j|^2-2|x_j|d_j+\frac{d_j^2}{1+\lambda}-2t_j),\quad (x,t)\in\mathbb{R}^{nm}\times\mathbb{R}^m,$$

and let $d^2=(d_1^2,\dots,d_m^2)\in\mathbb{R}^m$, then $H(x,-\lambda d^2/(2(1+\lambda))=H(x)$. Also

$$\begin{aligned}&\lambda|x|^2+(1+\lambda)|t|^2+H(x,t)\\&=(1+\lambda)\left\{\sum_{j=1}^m\left[\left(|x_j|-\frac{d_j}{1+\lambda}\right)^2+(t_j-\frac{1}{1+\lambda})^2\right]\right\}-\frac{m}{(1+\lambda)}\end{aligned}$$

$$= (1+\lambda)\left[\sum_{j=1}^{m} \operatorname{dist}^2((x_j,t_j), \frac{1}{1+\lambda}S_j^*)\right] - \frac{m}{(1+\lambda)}$$
$$= (1+\lambda)\operatorname{dist}^2\left((x,t),\ \frac{1}{1+\lambda}T^*\right) - \frac{m}{(1+\lambda)},$$

where $S_j^* = \{(y,1) \in \mathbb{R}^{n+1},\ y \in S_{d_j}$ is the $(n-1)$-dimensional sphere in $\mathbb{R}^{n+1}$ centred at $(0,1)$ with radius d_j while if we write $\mathbb{R}^{mn} \times \mathbb{R}^m = \{((x_1,t_1),\dots,(x_m,t_m)),\ x_j \in \mathbb{R}^n,\ t_j \in \mathbb{R},\ j=1,\dots,m\}$, then $T^* = \{(x,t) \in \mathbb{R}^{mn} \times \mathbb{R}^m,\ (x_i,t_i) \in S_j^*,\ j = 1,\dots,m\}$. Now we define the convex function $g_\lambda(x,t) = \lambda|x|^2 + (1+\lambda)|t|^2$ for $(x,t) \in \mathbb{R}^{mn} \times \mathbb{R}^m$, then

$$g_\lambda(x,t) + H_\lambda(x,t) = (1+\lambda)\operatorname{dist}^2\left((x,t),\ \frac{1}{1+\lambda}T^*\right) - \frac{m}{(1+\lambda)},$$

hence

$$C[g_\lambda(x,t) + H_\lambda(x,t)] = (1+\lambda)\operatorname{dist}^2\left((x,t),\ \frac{1}{1+\lambda}C(T^*)\right) - \frac{m}{(1+\lambda)}, \tag{41}$$

where $C(T^*)$ is the convex hull of T^*, so that

$$C^l_{g_\lambda}(H_\lambda(x,t)) = (1+\lambda)\operatorname{dist}^2\left((x,t),\ \frac{1}{1+\lambda}C(T^*)\right) - \frac{m}{(1+\lambda)} - g_\lambda(x,t), \quad (x,t) \in \mathbb{R}^{mn} \times \mathbb{R}^m. \tag{42}$$

We have

Theorem 3.1. *For $x \in \mathbb{R}^{mn}$, the family of functions*

$$\begin{aligned} F_\lambda(x) &= C^l_{g_\lambda}\left[H_\lambda\left(x, -\frac{\lambda d^2}{2(1+\lambda)}\right)\right] \\ &= (1+\lambda)\operatorname{dist}^2\left(\left(x,\ -\frac{\lambda d^2}{2(1+\lambda)}\right),\ \frac{1}{1+\lambda}C(T^*)\right) \\ &\quad - \frac{m}{(1+\lambda)} - \lambda|x|^2 - \frac{\lambda^2|d^2|}{4(1+\lambda)}, \end{aligned} \tag{43}$$

defined by evaluating $C^l_{g_\lambda}(H_\lambda(x,t))$ at $t = -\lambda d^2/(2(1+\lambda))$ for $\lambda > 0$ is a $C^{1,1}$ approximation of $H(x)$ as $\lambda \to +\infty$ with the estimates

$$0 \le H(x) - F_\lambda(x) \le \frac{\operatorname{diam}^2(T^*)}{1+\lambda}, \quad |DF_\lambda(x)| \le 2|T^*|, \quad x \in \mathbb{R}^n.$$

Let us apply the quadratic lower transforms to some one-dimensional examples.

The widely used fourth-order double-well function $f(x) = (x^2-1)^2$ vanishes exactly at 1 and -1. The advantage of such a function in modelling a double-well structure is that it is a fourth-order polynomial hence is smooth. The disadvantage of such a choice is that it is super-quadratic hence is more difficult to control. The natural geometric choice of the double-well function $g(x) = \operatorname{dist}^2(x, \{-1,1\}) = \min\{|x-1|^2,\ |x+1|^2\}$ is not smooth, therefore it cannot be used directly. Now we take the quadratic lower transform on $g(x)$ to obtain a $C^{1,1}$ approximation.

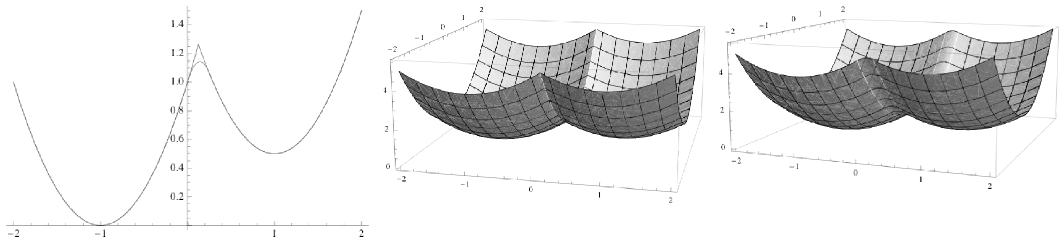

FIGURE 4. (Example 3.2) Left: $f(x)$ and $C^l_{2,\lambda}(f(x))$, middle: $h(x,y)$ and right: $C^l_{2,\lambda}(h(x,y))$ with $\alpha = 1/2$, $\lambda = 7$.

Example 3.2. Let $g(x) = \text{dist}^2(x, \{-1,1\}) = \min\{|x-1|^2, |x+1|^2\}$, $x \in \mathbb{R}$, then for $\lambda > 0$,

$$C^l_{2,\lambda}(g(x)) = \begin{cases} \dfrac{\lambda}{1+\lambda} - \lambda|x|^2, & |x| \le \dfrac{1}{1+\lambda}, \\ g(x), & |x| \ge \dfrac{1}{1+\lambda}. \end{cases}$$

We see that the lower transform simply smoothens out the corner and make the resulting function $C^{1,1}$. Clearly $C^l_{2,\lambda}(g(x)) \to g(x)$ uniformly as $\lambda \to +\infty$, and $[C^l_{2,\lambda}(g(x))]' \to g'(x)$ except at 0.

We would also like to present a slightly more general squared distance-like function and its lower transform. Consider $f(x) = \min\{(x+1)^2, (x-1)^2 + \alpha\}$ with $0 \le \alpha < 4$ and the function $h(x,y) = f(x) + y^2$ of two variables. We have $h(x,y) = \min\{|(x,y)-(-1,0)|^2, |(x,y)-(1,0)|^2 + \alpha\}$. We have

$$C^l_{2,\lambda}(f(x)) = \begin{cases} \dfrac{\alpha(\lambda+1)}{2}x - (\lambda+1)\left(\dfrac{\alpha}{4} - \dfrac{1}{\lambda+1}\right)^2 + 1 - \lambda x^2, \\ \qquad\qquad x \in \left[\dfrac{\alpha}{4} - \dfrac{1}{\lambda+1}, \dfrac{\alpha}{4} + \dfrac{1}{\lambda+1}\right], \\ f(x), \qquad \text{otherwise}. \end{cases}$$

It is also easy to see that $C^l_{2,\lambda}(h(x,y)) = C^l_{2,\lambda}(f(x)) + y^2$. In Figure 4 we plot the graphs of $f(x)$ and $C^l_{2,\lambda}(f(x))$ in the same picture with $\alpha = 1/2$, $\lambda = 7$. We also show the three dimensional graph of $h(x,y)$ and $C^l_{2,\lambda}(h(x,y))$.

Example 3.3. Now we consider a genuine two dimensional function. Let $K \subset \mathbb{R}^2$ be the four point set $K = \{(-1,1),,(1,1),(1,-1),(-1,-1)\}$ and let $f(x,y) =$

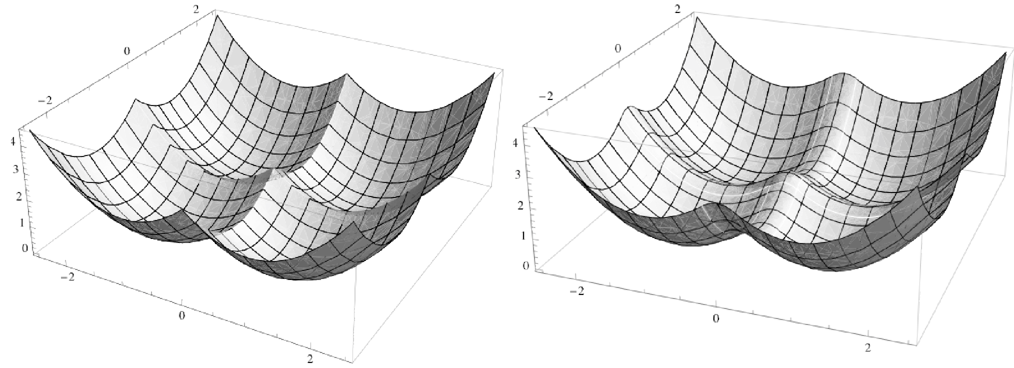

FIGURE 5. (Example 3.3) Left: $f(x,y)$, right: $C^l_{2,\lambda}(f(x,y))$ with $\lambda = 7$.

$\mathrm{dist}^2((x,y),K)$. For $\lambda > 0$,

$$C^l_{2,\lambda}f(x,y) = C[\mathrm{dist}^2((x,y),K) + \lambda|(x,y)|^2] - \lambda|(x,y)|^2$$

$$= \begin{cases} 2 - \dfrac{2}{1+\lambda} - \lambda(x^2+y^2), & -\dfrac{1}{1+\lambda} \le x,\ y \le \dfrac{1}{1+\lambda}, \\ f(x,y), & |x| \ge \dfrac{1}{1+\lambda}, \quad |y| \ge \dfrac{1}{1+\lambda}, \\ (1+\lambda)\left(x - \dfrac{1}{1+\lambda}\right)^2 + 2\left(1 - \dfrac{1}{1+\lambda}\right) - \lambda(x^2+y^2), & \\ & x \ge \dfrac{1}{1+\lambda},\ -\dfrac{1}{1+\lambda} \le y \le \dfrac{1}{1+\lambda}, \\ (1+\lambda)\left(x + \dfrac{1}{1+\lambda}\right)^2 + 2\left(1 - \dfrac{1}{1+\lambda}\right) - \lambda(x^2+y^2), & \\ & x \le -\dfrac{1}{1+\lambda},\ -\dfrac{1}{1+\lambda} \le y \le \dfrac{1}{1+\lambda}, \\ (1+\lambda)\left(y - \dfrac{1}{1+\lambda}\right)^2 + 2\left(1 - \dfrac{1}{1+\lambda}\right) - \lambda(x^2+y^2), & \\ & y \ge \dfrac{1}{1+\lambda},\ -\dfrac{1}{1+\lambda} \le x \le \dfrac{1}{1+\lambda}, \\ (1+\lambda)\left(y + \dfrac{1}{1+\lambda}\right)^2 + 2\left(1 - \dfrac{1}{1+\lambda}\right) - \lambda(x^2+y^2), & \\ & y \le -\dfrac{1}{1+\lambda},\ -\dfrac{1}{1+\lambda} \le x \le \dfrac{1}{1+\lambda}. \end{cases}$$

In Figure 5, we plot the graphs of $f(x,y)$ and $C^l_{2,\lambda}(f(x,y))$ for $\lambda = 4$ in the domain $|x| \le 2.5$, $|y| \le 2.5$.

Example 3.4. Consider the distance function

$$f(x) = \mathrm{dist}(x, \{-1,\ 1\}) = \min\{|x-1|,\ |x+1|\} = ||x|-1|, \qquad x \in \mathbb{R}.$$

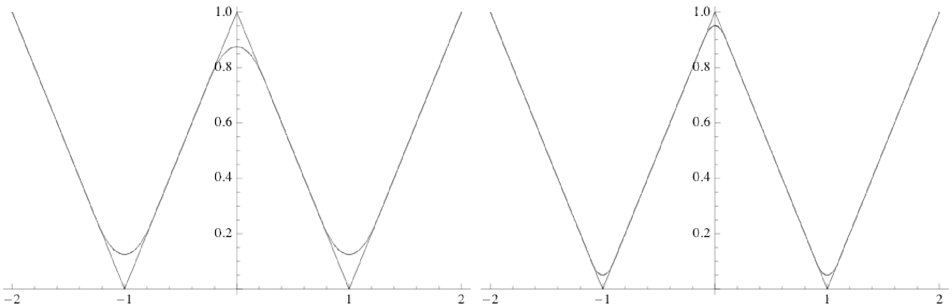

FIGURE 6. (Example 3.4) $f(x)$ and $C^l_{2,\lambda}(f(x))$. Left: $\lambda = \mu = 2$, right: $\lambda = \mu = 5$.

We have

$$C^l_{2,\lambda}(f(x)) = \begin{cases} 1 - \dfrac{1}{4\lambda} - \lambda x^2, & |x| \le \dfrac{1}{2\lambda}, \\ f(x), & |x| \ge \dfrac{1}{2\lambda}; \quad \left(\lambda \ge \dfrac{1}{2}\right). \end{cases}$$

$$C^u_{2,\mu}(f(x)) = \begin{cases} \mu(|x|-1)^2 + \dfrac{1}{4\mu}, & ||x|-1| \le \dfrac{1}{2\mu}, \\ f(x), & \text{otherwise}; \quad \left(\mu \ge \dfrac{1}{2}\right); \end{cases}$$

If we apply the lower transform first followed by an upper transform for $\lambda > 1$ and $\mu > 1$, we have

$$C^u_{2,\mu}[C^l_{2,\lambda}(\text{dist}(x, \{-1, 1\})] = \begin{cases} 1 - \dfrac{1}{4\lambda} - \lambda x^2, & |x| \le \dfrac{1}{2\lambda}, \\ \mu(|x|-1)^2 + \dfrac{1}{4\mu}, & ||x|-1| \le \dfrac{1}{2\mu}, \\ \text{dist}(x, \{-1, 1\}), & \text{otherwise.} \end{cases}$$

Due to the locality property of quadratic transforms, we see that in this example $C^u_{2,\mu}[C^l_{2,\lambda}(f(x)] = C^l_{2,\mu}[C^u_{2,\lambda}(f(x)]$ for large $\lambda > 0$ and $\mu > 0$ because the lower and upper transforms act on different part of the graph respectively, upper near non-smooth local convex points -1 and 1 while lower near the concave point 0. In Figure 6, we compare two sets of graphs between $f(x)$ and $C^l_{2,\lambda}(f(x))$ with $\lambda = \mu = 2$ and $\lambda = \mu = 5$ respectively.

Example 3.5. Example 3.4 can be easily extended to a two dimensional one. Consider

$$\begin{aligned} g(x, y) &= \text{dist}(x, \{-1, 1\}) + \text{dist}(y, \{-1, 1\}) \\ &= \min\{|x-1|, |x+1|\} + \min\{|x-1|, |x+1|\} \\ &= ||x|-1| + ||y|-1|. \end{aligned}$$

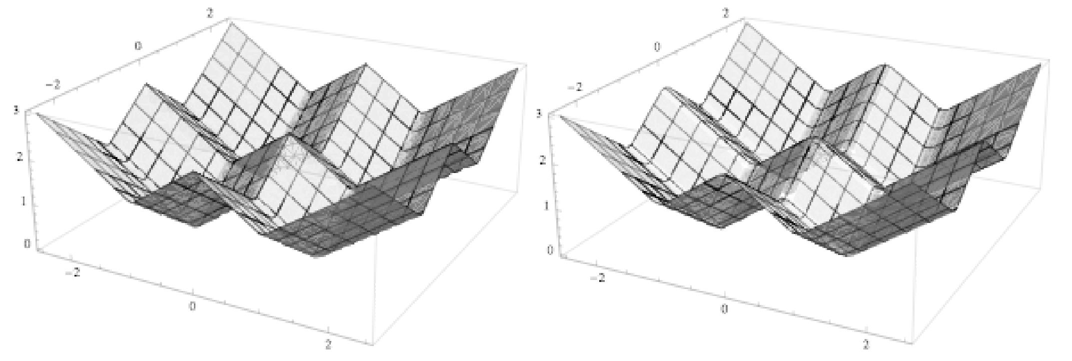

FIGURE 7. (Example 3.5) Left: graph of $g(x,y)$ and right: $C^u_{2,\lambda}[C^l_{2,\lambda}(g(x,y))]$ with $\lambda = \mu = 5$.

Due to the special form $g(x,y) = f(x) + f(y)$, we see, by the definition of convex envelope that

$$\begin{aligned} C[g(x,y) + \lambda(|x|^2 + |y|^2)] &= C[(f(x) + \lambda|x|^2) + (f(y) + \lambda|y|^2)] \\ &= C[(x) + \lambda|x|^2] + C[f(y) + \lambda|y|^2]. \end{aligned}$$

Therefore

$$C^l_{2,\lambda}[g(x,y)] = C^l_{2,\lambda}[f(x)] + C^l_{2,\lambda}[f(y)],$$

hence

$$C^u_{2,\tau}[C^l_{2,\lambda}[g(x,y)]] = C^u_{2,\tau}[C^l_{2,\lambda}[f(x)]] + C^u_{2,\tau}[C^l_{2,\lambda}[f(y)]].$$

Thus for $\lambda > 0$ and $\tau > 0$ small, $C^u_{2,\tau}[C^l_{2,\lambda}[g(x,y)]] = C^l_{2,\lambda}[C^u_{2,\tau}[g(x,y)]]$ and it defines a tight $C^{1,1}$ approximation for $g(x,y)$. Figure 7 shows how $C^u_{2,\tau}[C^l_{2,\lambda}[g(x,y)]]$ smoothes $g(x,y)$ near the edges and corners of the graph of $g(x,y)$. Observe that for $\lambda = \mu = 5$, the two graphs are almost the same except at the edges and corners where $C^l_{2,\lambda}[C^u_{2,\tau}[g(x,y)]]$ smoothes them out.

Acknowledgement. This work is partially supported by Wales Institute for Mathematical and Computing Sciences and China National Science Foundation through grant 40771094. I also wish to thank Dr Ian M. Davis for his helpful suggestions.

REFERENCES

[1] F. Alizadeh, D. Goldfarb, *Second-order cone programming,* Math. Program. Series B **95** (2003), 3–51.

[2] L. T. H. An, *Solving large scale modecular distance geometry problems by a smoothing technique via the Gaussian transform and D.C. programming,* J. Global Optim. **27** (2003), 375–397.

[3] P. Armand, *A quasi-Newton penalty barrier method for convex minimization problems*, Comp. Optim. Appl. **26** (2003), 5–34.

[4] A. Auslander, *Penalty and barrier methods: a unified framework* SIAM J. Optim. **10** (1999), 211–230.

[5] D. P. Bertsekas, *Constrained optimization and Lagrangian multiplier methods*, Academic Press, 1982.

[6] S. I. Birbil, S.-C. Fang, J. B. G. Frenk, S. Zhang, *Recursive approximation of the high dimensional* max *function,* Oper. Res. Lett. **33** (2005), 450–458.

[7] H. Blum, *A transformation for extracting new descriptors of shape,* Prop. Symp. Models for the Perception of Speech and Visual Form (W. W. Dunn ed.), MIT Press (1967), 362–380.
[8] A. Brøndsted, *Introduction to convex ploytopes*, Springer, 1982.
[9] A. Ben-Tal, M. Teboulle, *A smoothing technique for nondifferentialble optimization problems,* in Optimization, Lect. Notes in Math **1405** Springer (1989), 1–11.
[10] S. Boyd, L. Vandernberghe, *Convex programming,* Cambridge Univ. Press, 2004.
[11] F. H. Clarke *Optimization and nonsmooth analysis,* J. Wiley, 1983.
[12] H. I. Choi, S. W. Choi, H. P. Moon, *Mathematical theory of medial axis transform* Pacific J. Math. **181** (1997), 57–88.
[13] C. Chen, O. L. Mangasarian, *Smoothing methods for convex inequalities and linear complementarity problems*, Math. Program. bf 71 (1995), 51–69.
[14] X. Chen, H. Qi, L. Qi, K.-K. Teo, *Smooth convex approximation to the maximum eigenvalue function* J. Global Optim. **30** (2004), 253–270.
[15] G. M. Crippen, T. F. Havel, *Distance geometry and molecular conformation,* John Wiley & Sons, 1988.
[16] F. Deutsch *Best approximation in inner product spaces*, Springer-Verlag, New York, 2001.
[17] S. Guillaume, A. Seeger *A higher-order smoothing technique for polyhedral convex functions: Geometry and probabilistic considerations* J. Convex Anal. **8** (2001), 109–126.
[18] T. F. Havel, *An evaluation of computational strategies for use in the determination of protein structure from distance geometry constraints obtained by nuclear magnetic resonance,* Prog. Biol. **56** (1991), 43–78.
[19] C. Kanzow *Some noninterior continuition method for linear complementarity problems* SIAM J. Matrix Anal. Appl. **17** (1996), 851–868.
[20] S. R. Lay, *Convex Sets and Their Applications* John Wiley & Sons, 1982.
[21] X. S. Li, *Aggregation function method for nonlinear programming* Science in China (Series A) **34** (1991), 1467–1473.
[22] R. F. Love, J. G. Morris, G. O. Wesolowsky, *Facilities location: Models and methods* North-Holland, 1998.
[23] C. Mantegazza, A. C. Mennacci, *Hamilton-Jacobi equations and distance functions on Riemannian manifolds*, Appl. Math. Optim. **47** (2002), 1–25.
[24] J. J. More, Z. Wu, *Global contiuition for distance geometry problem,* SIAM J. Optim. **8** (1997), 814–836.
[25] Yu. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program. Ser. A **103** (2005), 127–152.
[26] J.-M. Peng, Z.-H. Lin, *A non-interior continuation method for generalized linear complementarity problems,* Math. Program. **86** (1999), 533–563.
[27] E. Polak, *Optimizations: Algorithms and consistent approximations*, Springer-Verlag, 1997.
[28] L. Qi, X. Chen, *A globally convergent successive approximation methods for non-smooth equations,* SIAM J. Control Optim. **33** (1995), 402–418.
[29] H.-D. Qi, L.-Z. Liao, *A smoothing Newton method for extended vertical linear complementarity problems*, SIAM J. Matrix Anal. Appl. **21** (1999), 45–66.
[30] H.-D. Qi, P. Tseng, *Analysis of piecewise smooth functions and almost smooth functions*, Preprint (2002).
[31] R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, 1966.
[32] R. L. Sheu, J. Y. Lin, *Solving continuous min-max problems by an iterative entropic regularization method*, J. Optim. Theory Appl. **121** (2004), 597–612.
[33] H. J. H. Tuenter, *The minimum L_2-distance projection onto the canonical simplex: A simple algorithm*, ALGO Research Quarterly, December, 2001.
[34] P. Tseng, D. P. Bertsekas, *On the convergence of the exponential multiplier method for convex programming*, Math. Program. **60** (1993), 682–684.
[35] H. Tang, L. Zhang, *A maximum entropy method for convex programming,* Chinese Science Bull. **39** (1994), 1–19.
[36] G. Voronoi, *Nouvelles applications des parametres continus a la theorie des formes quadratiques*, J. Reine Angew. Math. **134** (1908), 198–287.

[37] F. E. Wolter, *Cut locus and medial axis in global shape interrogation and representation*, MIT, Dept. Ocean Engineering, Design Laboratory Memorandum Issue 92-2, 1993.

[38] S. Xu, *Smoothing method for minimax problems,* Comp. Optim. Appl. **20** (2001), 267–279.

[39] K. Zhang, *On some quasiconvex functions with linear growth,* J. Convex Anal. **5** (1998), 133–146.

[40] K. Zhang, *A two-well structure and intrinsic mountain pass points*, Calc. Var. PDEs **13** (2001), 231–264.

[41] K. Zhang, *Maximal extension for linear spaces of real matrices with large rank*, Proc. Royal Soc. Edinb. **131A** (2001), 1481–1491.

[42] K. Zhang, *An elementary derivation of the generalized Kohn-Strang relaxation formulae*, J. Convex Anal. **9** (2002), 269–285.

[43] K. Zhang *On conditions for equality of relaxations in the calculus of variations* J. Nonl. Convex Anal. **3** (2002), 145–154.

[44] K. Zhang, *On separation of gradient Young measures*, Calc. Var. PDEs **17** (2003) 85–103.

[45] K. Zhang, *Compensated convexity and its applications,* Anal. nonlin. H. Poincare Inst. **25** (2008), 743–771.

Manuscript received August 13, 2008
revised October 9, 2008

Kewei Zhang
Department of Mathematics, Swansea University, Singleton Park, Swansea, SA2 8PP, UK
E-mail address: k.zhang@swansea.ac.uk