MATHEMATICS OF OPERATIONS RESEARCH
Vol. 30, No. 4, November 2005, pp. 785–799
ISSN 0364-765X | EISSN 1526-5471 | 05 | 3004 | 0785

DOI 10.1287/moor.1050.0163

Estimating Tangent and Normal Cones Without Calculus

W. L. Hare
Advanced Optimization Laboratory, McMaster University, Hamilton, Ontario, Canada L8S 4K1,
whare@cecm.sfu.ca, http://optlab.mcmaster.ca/∼whare/

A. S. Lewis
School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853,
aslewis@orie.cornell.edu, http://www.orie.cornell.edu/∼aslewis

We study simple estimates of the tangent and normal cones to a set with a nonsmooth boundary. We seek estimates that are robust relative to the base point and that depend only on collections of nearby points known to lie inside and outside of the set.

Key words: nonsmooth optimization; variational analysis; normal cone; proximal normal; tangent cone; Clarke regular
MSC2000 subject classification: Primary: 49J52, 65K10; secondary: 90C46, 58C06
OR/MS subject classification: Primary: programming, nondifferentiable; secondary: mathematics, convexity
History: Received April 30, 2004; revised December 8, 2004.

1. The tangent cone. We can estimate the directional derivative and gradient of a smooth function quickly and easily using finite difference formulas. While rather inaccurate, such estimates have some appeal, needing neither calculus rules nor even a closed-form expression for the function. In the variational geometry of sets, the role of derivatives and gradients are played by the cones of "tangent" and "normal" vectors. We study here how we might estimate these cones, without any prior knowledge of the structure of the set (like convexity, for example), and without recourse either to the calculus rules of nonsmooth analysis or even to an analytic description of the set. We rely instead only on the most primitive description of the set, namely a membership *oracle*—an algorithm that decides whether or not any given point belongs to the set.

Our interest is both philosophical and practical. Philosophically, are the tangent and normal cones in any sense computable from a primitive check on set membership? Practically, could we design a subroutine for estimating these cones without requiring any structural knowledge of the set from the user?

Like differentiation, the idea of the tangent cone involves a limit, but one involving sets. We therefore present some notions of set convergence, before defining the tangent cone. Comprehensive presentations of variational analysis and nonsmooth optimization can be found in Clarke et al. [4] or Rockafellar and Wets [10]. We follow the notation and terminology of the latter, unless otherwise stated.

Definition 1.1. Given a family of sets $D_\epsilon \subset \mathbf{R}^n$ indexed by $\epsilon > 0$, the *outer* and *inner limits* are defined, respectively, by

$$\limsup_{\epsilon\downarrow 0} D_\epsilon = \{x\colon \exists\, x_k \to x,\ \exists\, \epsilon_k \downarrow 0 \text{ with } x_k \in D_{\epsilon_k}\},$$

$$\liminf_{\epsilon\downarrow 0} D_\epsilon = \{x\colon \forall \epsilon_k \downarrow 0,\ \exists\, x_k \to x \text{ with } x_k \in D_{\epsilon_k}\}.$$

If these two sets both equal a set $D \subset \mathbf{R}^n$, then we say that D_ϵ *converges to* D, and we write $\lim_{\epsilon\downarrow 0} D_\epsilon = D$.
Notice the inner limit is always contained in the outer limit.

The definition of the tangent cone is rather analogous to the familiar definition of the directional derivative of a real function f at a point $\bar{x}$ in a direction d:

$$f'(\bar{x}; d) = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}(f(\bar{x} + \epsilon d) - f(\bar{x})).$$

Definition 1.2. The *tangent cone* to a set $S \subset \mathbf{R}^n$ at a point $\bar{x} \in S$ is the set

$$T_S(\bar{x}) = \limsup_{\epsilon \downarrow 0} \frac{1}{\epsilon}(S - \bar{x}).$$

This cone is sometimes called the "Bouligand" or "contingent" cone.

Consider a set $S \subset \mathbf{R}^n$ with $0 \in S$, and suppose we wish to estimate $T_S(0)$. A natural approach is to test some set $A \subset \mathbf{R}^n$ of points near 0 for membership in S, and then to estimate

$$T_S(0) \approx \mathbf{R}_+(S \cap A). \tag{1}$$

In practice, the set A would be finite.

Precisely how we choose the set A is not an issue we pursue here. One possible approach is to generate it by sampling independent, uniformly distributed points in some neighborhood of 0. An analogous idea for estimating generalized gradients is described in Burke et al. [1], and a nonsmooth optimization algorithm based on this idea performs quite well in practice (see Burke et al. [2]).

We can formalize what we mean by A consisting of "points near 0" by considering a sequence of sets

$$A_k \subset \rho_k B,$$

where $B \subset \mathbf{R}^n$ denotes the closed Euclidean unit ball and the scalars $\rho_k \in \mathbf{R}_+$ approach 0. In §3 we study under what conditions the set $\mathbf{R}_+(S \cap A_k)$ approaches $T_S(0)$. The notions of convergence for sequences of sets are completely analogous to Definition 1.1, so we do not reproduce them (see Rockafellar and Wets [10, Definition 4.1]). It is easy to check

$$\limsup_{k \to \infty} \mathbf{R}_+(S \cap A_k) \subset T_S(0). \tag{2}$$

We cannot expect a stronger result without assuming more about the sets A_k. To capture all possible tangent vectors, the least we could reasonably impose is that any direction in $\mathbf{R}^n$ is a limit of directions generated by A_k. In other words, we make the assumption

$$\lim_k A_k = \{0\} \qquad \text{and} \qquad \lim_k \mathbf{R}_+ A_k = \mathbf{R}^n. \tag{3}$$

This condition is implementable: For example, we could fix any dense sequence $\{x_1, x_2, x_3, \ldots\}$ in the ball B and then define

$$A_k = \frac{1}{k}\{x_1, x_2, \ldots, x_k\}.$$

Unfortunately, Assumption (3) does not alone suffice to guarantee any more than the one-sided estimate (2) of the tangent cone. Consider for example, in $\mathbf{R}^2$, the set S consisting of the first coordinate axis, and the sets

$$A_k = \{(u_1, u_2)\colon\ k\|u\|_2 \le 1,\ \ k|u_2| \ge |u_1|\}.$$

A little thought suggests that for this approach to succeed without asking too much of the sets A_k, the set S must have reasonable "interiority" properties. In §3 we consider the case where S is the epigraph of a Lipschitz function $f\colon \mathbf{R}^{n-1} \to \mathbf{R}$:

$$S = \operatorname{epi} f = \{(y, s) \in \mathbf{R}^{n-1} \times \mathbf{R}\colon\ f(y) \le s\}$$

(where $f(0) = 0$). In this case, assumption (3) guarantees our desired estimate

$$\lim_k \mathbf{R}_+(S \cap A_k) = T_S(0),$$

providing S is also "Clarke regular" at 0. We discuss this crucial property in the next section.

While simple and natural, the approach we have sketched is not entirely satisfactory. First, in our estimate, we simply discard those points in the set A_k that we discover lie outside the set S. Can we not use the geometric information inherent in these points more strongly? More serious, however, are the flaws in our tangent cone estimate if we try to use it to estimate its dual object, the "normal cone." We discuss these difficulties, and other estimates of the normal cone, in the next section.

2. The normal cone. Fundamental to the variational geometry of a closed set $S \subset \mathbf{R}^n$ at a point $\bar{x} \in S$ is the idea of the normal cone $N_S(\bar{x})$ defined below. For example, we can characterize the boundary of S elegantly by

$$\bar{x} \in \text{ boundary of } S \Leftrightarrow N_S(\bar{x}) \neq \{0\}. \tag{4}$$

The normal cone is crucial in the development of optimality conditions. For example, if $\bar{x}$ is a local maximizer of a smooth function f over S, then $\nabla f(\bar{x}) \in N_S(\bar{x})$. In this section we parallel our investigation of the tangent cone in the previous section, studying estimates of the normal cone based on the membership in S of points lying in some given set near $\bar{x}$.

The polar of a set $K \subset \mathbf{R}^n$ is the closed convex cone

$$K^* = \{y \in \mathbf{R}^n \colon \langle x, y\rangle \leq 0 \ \forall x \in K\}.$$

When the set S at the point $\bar{x} \in S$ has the property of Clarke regularity mentioned at the end of the previous section, the normal and tangent cones are mutually polar. In this case, following the approach of the previous section and estimating the tangent cone by formula (1) implies the estimate

$$N_S(0) \approx (S \cap A)^*. \tag{5}$$

However, simple examples show this estimate is hopeless. In $\mathbf{R}^2$, consider the set $S = \{(u, v)\colon v \geq -u^2\}$, and the set $A = \epsilon B$ for any scalar $\epsilon > 0$. The estimate above gives $N_S(0) \approx \{0\}$, whereas in fact $N_S(0) = \mathbf{R}_+(0, -1)$.

Even when the set S is convex, the estimate (5) seems unsatisfactory, due to a certain lack of "robustness." Imagine estimating the normal cone to the right halfplane in $\mathbf{R}^2$, for example, at a point $\bar{x}$ in the interior of the halfplane but close to the boundary. If the set A we use is of a scale larger than the distance to the boundary, the oracle will typically detect points outside the halfplane, correctly indicating that $\bar{x}$ is near the boundary. However, this information is often lost in the estimate (5); instead, a little thought shows a typical estimate is $N_S(\bar{x}) \approx \{0\}$.

These failures suggest trying to estimate the normal cone $N_S(\bar{x})$ using points near $\bar{x}$ both inside *and* outside the set S. To fix some notation, suppose $0 \in S$ and that we know two subsets:

$$A_{\text{out}} \subset S^c \cap \epsilon B \qquad \text{and} \qquad A_{\text{in}} \subset S \cap \epsilon B, \tag{6}$$

where S^c denotes the complement of S, ϵ is a small positive scalar, and $0 \in A_{\text{in}}$. For example, A_{out} and A_{in} might be finite sets generated randomly and distinguished by the oracle. On this basis, how might we estimate the normal cone $N_S(0)$?

To approach this question, we first need the definition of the normal cone. Both from an historical perspective (beginning with Mordukhovich [8], Kruger and Mordukhovich [7], and Ioffe [6]), and for geometric transparency (see Clarke et al. [4], for example), it is attractive to consider the normal cone as the cone of "limiting proximal normals," defined below. Although Rockafellar and Wets [10] emphasize a more analytic approach, for our present purposes, proximal normals are an illuminating tool. Given any point $x \in \mathbf{R}^n$, we denote the set of closest points to x in S by $P_S(x)$.

Definition 2.1. Given any set $S \subset \mathbf{R}^n$, the *proximal normal cone* to S at a point $y \in S$ is the cone

$$N_S^P(y) = \{\lambda(x - y)\colon \lambda \geq 0, y \in P_S(x)\}.$$

The proximal normal cone thus consists of vectors (called "proximal normals") in the direction of points that project back onto y.

If the set S is closed, we can now define the normal cone by closing the graph of the multifunction $N_S^P(\cdot)$, as follows (see Clarke et al. [4, Theorem 6.1]).

Definition 2.2. Given any set $S \subset \mathbf{R}^n$, the *normal cone* to S at a point $y \in S$ is the cone

$$N_S(y) = \left\{w\colon \exists\, y_k \in S,\ \exists\, w_k \in N_S^P(y_y) \text{ with } y_k \to y, w_k \to w\right\}.$$

This construction motivates the term "limiting proximal normal cone." The coincidence of this cone with various equivalent constructions is discussed in Rockafellar and Wets [10].

Given the central role of proximal normals, a first try at using the inclusions (6) might be to estimate the normal cone $N_S(0)$ by the set

$$E_\epsilon = \mathbf{R}_+ \bigcup_{x \in A_{\text{out}}} (x - P_{A_{\text{in}}}(x)) \tag{7}$$

As we apply the oracle to more and more points, we might consider an ideal, limiting situation where

$$A_{\text{out}} = S^c \cap \mu\epsilon B \qquad \text{and} \qquad A_{\text{in}} = S \cap \gamma\epsilon B, \tag{8}$$

for some constants $\mu, \gamma > 0$.

For such an approach to be promising, it seems reasonable to demand that the estimate (7) converges to the normal cone $N_S(0)$ as $\epsilon \downarrow 0$.

In §4, we prove convergence results for sets of the form (7). For example, we show that the set

$$N_\epsilon = \mathbf{R}_+ \bigcup_{x \in 2\epsilon B} (x - P_{S \cap 5\epsilon B}(x)) \tag{9}$$

converges to $N_S(0)$ as $\epsilon \downarrow 0$.

Unfortunately, this result seems useless in practice. Any attempt to estimate normal cones by randomly generating independent nearby points, using the oracle to construct subsets A_{out} and A_{in} approximating those in equations (8), and then using the estimate (7), seems hopelessly inaccurate in simple experiments. Loosely speaking, the difficulty is caused by nearby points $x \in A_{\text{out}}$ and $y \in A_{\text{in}}$ generating short "displacements" $x - y$ that do not reasonably approximate any proximal normal.

We can avoid this unstable behavior by ignoring displacements that are too short. For example, again in the ideal, limiting case, we could consider the estimate

$$N_\epsilon = \{0\} \cup \mathbf{R}_+ \bigcup_{x \in 2\epsilon B} ((x - P_{S \cap 5\epsilon B}(x)) \cap \epsilon \operatorname{cl}(B^c)). \tag{10}$$

In general, this approach fails. Consider, for example, the closed subset of $\mathbf{R}$

$$S = \{0\} \cup \left\{\pm\frac{1}{n}\colon n \in \mathbf{N}\right\}.$$

For all small $\epsilon > 0$, it is easy to check that formula (10) gives $N_\epsilon = \{0\}$, and yet $N_S(0) = \mathbf{R}$.

We can remedy this by adding a "regularity" condition. We describe this condition in the form presented in Poliquin et al. [9, Theorem 1.3].

Definition 2.3. A set $S \subset \mathbf{R}^n$ is *prox-regular* at a point $\bar{x} \in S$ if S is locally closed at $\bar{x}$ and the projection $P_S(x)$ is a singleton for all points $x \in \mathbf{R}^n$ near $\bar{x}$.

Prox-regularity is a rather strong assumption, but one that holds, for example, for all convex sets, and for all sets defined by a finite number of smooth constraints satisfying a reasonable constraint qualification (see Rockafellar and Wets [10, p. 612]). We show in §4 that the sets N_ϵ defined by (10) do indeed converge to the normal cone $N_S(0)$ as $\epsilon \downarrow 0$, providing S is prox-regular at 0.

A weaker condition than prox-regularity (see Poliquin et al. [9]), and one central to variational analysis, is the idea mentioned above of Clarke regularity. We follow Clarke's original definition (see Clarke [3] and also Rockafellar and Wets [10, Corollary 6.29]).

Definition 2.4. A set $S \subset \mathbf{R}^n$ is *Clarke regular* at a point $\bar{x} \in S$ if S is locally closed at $\bar{x}$ and the tangent and normal cones $T_S(\bar{x})$ and $N_S(\bar{x})$ are mutually polar.

When the set S is Clarke regular at $\bar{x}$, the limit defining the tangent cone (Definition 1.2) is well behaved:

$$T_S(\bar{x}) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon}(S - \bar{x}) \tag{11}$$

(see Rockafellar and Wets [10, Corollary 6.30]). This suggests a refinement in our approach. If S is Clarke regular at 0, then for small $\epsilon > 0$ we know

$$S \cap \epsilon B \approx T_S(0) \cap \epsilon B,$$

in a sense made precise by equation (11). Since the right hand side is convex, a more useful estimate of $S \cap \epsilon B$ than the approximation A_{in} from (6) may be its convex hull $\operatorname{conv} A_{\mathrm{in}}$. We pursue this approach in §5.

The assumption that the set S is Clarke regular at 0 is very significant for estimates of the normal cone $N_S(0)$. Since this cone is closed and convex, we may find it natural to restrict attention to estimates $\widetilde{N}_\epsilon$ that are themselves closed convex cones. Particularly when we are using normal cone approximations to estimate the tangent cone via polarity, this restriction involves no essential loss of generality. To see this, recall that the bipolar K^{**} of a set $K \subset \mathbf{R}^n$ is simply the closed convex cone generated by K, and furthermore, convergence of a sequence of closed convex cones is equivalent to convergence of the polars, by Rockafellar and Wets [10, Corollary 11.35]. Hence, we deduce $\widetilde{N}_\epsilon^* \to T_S(0)$ if and only if $\widetilde{N}_\epsilon^{**} \to N_S(0)$.

Suppose we have approximations

$$A_{\mathrm{out}} \approx S^c \cap \epsilon B \qquad \text{and} \qquad A_{\mathrm{in}} \approx S \cap \epsilon B.$$

Following our discussion above, we might estimate the normal cone $N_S(0)$ by the set

$$\widetilde{N}_\epsilon = \operatorname{conv}\left(\{0\} \cup \mathbf{R}_+\left((\mu\epsilon \operatorname{cl}(B^c)) \cap \bigcup_{x \in A_{\mathrm{out}}} \left(x - P_{\operatorname{conv} A_{\mathrm{in}}}(x)\right)\right)\right). \tag{12}$$

Keeping in mind the instability caused by small displacements that we noted before, we fix the constant $\mu \in (0, 1)$. We give an example in §5 to show the importance of this restriction.

Unfortunately, even in the ideal case where

$$A_{\mathrm{out}} = S^c \cap \epsilon B \qquad \text{and} \qquad A_{\mathrm{in}} = S \cap \epsilon B, \tag{13}$$

and the set S is Clarke regular at 0, the sets $\widetilde{N}_\epsilon$ may not converge to the normal cone $N_S(0)$. Consider, for example, the set $S = \{x \in \mathbf{R}^2\colon x_2 = x_1^2\}$. It is easy to check $\widetilde{N}_\epsilon = \mathbf{R}^2$ for all $\epsilon > 0$, and yet $N_S(0) = \{x\colon x_1 = 0\}$.

We introduce one further condition (see Rockafellar and Wets [10, p. 385, Example 9.42, Theorem 6.28]), to ensure that the set $\widetilde{N}_\epsilon$ is a reasonable estimate of the normal cone $N_S(0)$.

Definition 2.5. A set $S \subset \mathbf{R}^n$ is *epi-Lipschitz* at a boundary point $\bar{x}$ if S is locally closed at $\bar{x}$ and $(N_S(\bar{x}))^*$ has nonempty interior.

Loosely speaking, a set is epi-Lipschitz when, in a suitable coordinate system, we can view it locally as the epigraph of a Lipschitz function. This condition holds, for example, for convex sets with nonempty interior, and for sets defined by smooth inequalities.

When the set S is Clarke regular and epi-Lipschitz at 0, in the ideal case (13), we prove in §5 that

$$\lim_{\epsilon \downarrow 0} \widetilde{N}_\epsilon = N_S(0).$$

This theoretical property motivates our interest in the estimate $\widetilde{N}_\epsilon$. For comparison, our earlier estimate N_ϵ in (10) involved an additional scaling constant in its definition, and needs the assumption of prox-regularity to guarantee convergence.

In practice, on the other hand, we can typically only consider finite sets A_{out} and A_{in}. In this case, identifying the estimate N_ϵ is straightforward, but the set $\widetilde{N}_\epsilon$ is also relatively easy to compute, by solving a second-order cone problem for each point in A_{out}. Simple computational experiments suggest that the convex hull operations in the definition of $\widetilde{N}_\epsilon$ enhance it over N_ϵ as an estimate of the normal cone.

3. Tangent estimates. We begin our more detailed development with a result justifying our first estimate (1) of the tangent cone.

THEOREM 3.1 (TANGENT ESTIMATION). *Consider a set $S \subset \mathbf{R}^n$, a point $\bar{x}$ in S, and a sequence of sets $A_k \subset \mathbf{R}^n$ (for $k = 1, 2, 3, \ldots$), satisfying*

$$\lim_{k\to\infty} A_k = \{0\}.$$

Then,

$$\limsup_{k\to\infty} \mathbf{R}_+((S - \bar{x}) \cap A_k) \subset T_S(\bar{x}). \tag{14}$$

Suppose furthermore that

$$\lim_{k\to\infty} \mathbf{R}_+ A_k = \mathbf{R}^n, \tag{15}$$

and that either $\bar{x}$ lies in the interior of S, or that it lies in the boundary and S is epi-Lipschitz there. Then,

$$\liminf_{k\to\infty} \mathbf{R}_+((S - \bar{x}) \cap A_k) \supset (N_S(\bar{x}))^*. \tag{16}$$

Hence, if S is also Clarke regular at $\bar{x}$, then

$$\lim_{k\to\infty} \mathbf{R}_+((S - \bar{x}) \cap A_k) = T_S(\bar{x}) \tag{17}$$

PROOF. We lose no generality in supposing $\bar{x} = 0$.

To prove the inclusion (14), consider any subsequence M of the natural numbers, and consider points $x_m \in S \cap A_m$ and scalars $\lambda_m \in \mathbf{R}_+$ (for $m \in M$), such that $\lambda_m x_m$ approaches some limit $d \in \mathbf{R}^n$ as $m \to \infty$ in M. We want to show $d \in T_S(0)$. This holds if $d = 0$, so suppose $d \neq 0$. As $m \to \infty$ in M, we have $x_m \to 0$ and $\lambda_m x_m \to d$, so $\lambda_m \to +\infty$. Since $x_m \in S$ for all m, the result follows.

Moving to inclusion (16), the assumptions on S amount to the condition that S is locally closed at 0 and the cone $(N_S(0))^*$ has nonempty interior. This cone is in fact the "regular tangent cone" $\widehat{T}_S(0)$ (see Rockafellar and Wets [10, Definition 6.35 and Theorem 6.28]). Consider any vector $d \in \operatorname{int} \widehat{T}_S(0)$. Using Rockafellar and Wets [10, Theorem 6.36], there exist scalars $\epsilon, \delta > 0$ such that

$$[0, \epsilon](d + \delta B) \subset S. \tag{18}$$

By assumption (15), there exist sequences of scalars $\lambda_k \geq 0$ and vectors $x_k \in A_k$ such that $\lambda_k x_k \to d$. Hence, we deduce

$$\lambda_k x_k \in d + \delta B \quad \text{for all large } k. \tag{19}$$

We next claim $x_k \in S$ for all large k. To see this, consider first the case $d = 0$. In this case, inclusion (18) shows $0 \in \operatorname{int} S$, so since $x_k \to 0$, the claim follows. On the other hand, if $d \neq 0$, then $\lambda_k \to +\infty$, so the relation (19) implies

$$x_k \in [0, \epsilon](d + \delta B) \quad \text{for all large } k.$$

Our claim now follows, again from inclusion (18).

Summarizing, we now have scalars $\lambda_k \in \mathbf{R}_+$ and vectors $x_k \in S \cap A_k$ satisfying $\lambda_k x_k \to d$. We have thus proved

$$\operatorname{int} \widehat{T}_S(0) \subset \liminf_{k\to\infty} \mathbf{R}_+(S \cap A_k).$$

Since the right-hand side is closed (by Rockafellar and Wets [10, Proposition 4.4]), we deduce

$$\operatorname{cl}\left(\operatorname{int} \widehat{T}_S(0)\right) \subset \liminf_{k\to\infty} \mathbf{R}_+(S \cap A_k). \tag{20}$$

As we observed above, our epi-Lipschitz assumption ensures that the regular tangent cone $\widehat{T}_S(0)$ has nonempty interior. This cone is closed and convex, and so equals the left-hand side of inclusion (20). This completes the proof of inclusion (16). Equation (17) now follows from our definition of Clarke regularity (Definition 2.4). □

We end this section with some examples illustrating the above result. First, notice that the inclusion (16) can be strict even for epi-Lipschitz sets. In $\mathbf{R}^2$, consider the set $S = \{(u, v)\colon\ v \geq -|u|\}$. A quick calculation shows $N_S(0, 0) = \{(u, v)\colon\ v = -|u|\}$, so $(N_S(0, 0))^* = \{(u, v)\colon v \geq |u|\}$. However, if we choose $A_k = k^{-1}B$, for example, then $\mathbf{R}_+(S \cap A_k) = S$ for all $k = 1, 2, 3, \ldots$, so the inclusion (16) is strict. Notice S is epi-Lipschitz at the origin. Equation (17) fails because S is not Clarke regular at the origin.

Inclusion (16) can fail if the set S is not epi-Lipschitz. In $\mathbf{R}^2$, consider the set $S = \{(u, v)\colon\ u = 0\}$, for which we have $(N_S(0, 0))^* = F$. If we choose $A_k = \{(u, v)\colon\ k|v| \leq k^2|u| \leq 1\}$, then $S \cap A_k = \{0\}$ for all $k = 1, 2, 3, \ldots$, so the inclusion (16) fails. Notice S is Clarke regular at the origin.

Our last example shows that the inclusion (14) can be strict even when the set S is epi-Lipschitz. Consider the Lipschitz function $f\colon \mathbf{R} \to \mathbf{R}$ defined by

$$f(u) = \begin{cases} 3(|u| - 2^{-m}) & \text{if } 2 \leq 2^{1+m}|u| \leq 3,\ \ m \in \mathbf{N} \\ -3(|u| - 2^{1-m}) & \text{if } 3 \leq 2^{1+m}|u| \leq 4,\ \ m \in \mathbf{N} \\ 0 & \text{if } u = 0. \end{cases}$$

Notice $f(\pm 2^{-m}) = 0$ and $f(\pm 3.2^{-m}) = 3.2^{-m}$ for each integer $m \in \mathbf{N}$. Let S be the epigraph of f, so S is certainly epi-Lipschitz at the origin. A quick check shows $T_S(0, 0) = \{(u, v)\colon v \geq 0\}$. Now let A_k denote the square with corners $(\pm 3.2^{-k}, \pm 3.2^{-k})$. This sequence of sets clearly satisfies $\lim_k A_k = \{0\}$, and furthermore $\mathbf{R}_+ A_k = \mathbf{R}^2$ for all $k \in \mathbf{N}$. However, since for each $k \in \mathbf{N}$,

$$S \cap A_k = \{(u, 3.2^{-k})\colon |u| \leq 3.2^{-k}\},$$

we deduce $\mathbf{R}_+(S \cap A_k) = \{(u, v)\colon v \geq |u|\}$. Hence, the inclusion (14) is strict. Equation (17) again fails because S is not Clarke regular at the origin.

4. Normal estimates without convexification. In this section we turn to estimates of the normal cone. We consider estimates based on sets of neighboring points inside and outside the set, and prove the convergence of normal cone estimates like (9) and (10), that involve no convex hulls.

Recall that a closed set $S \subset \mathbf{R}^n$ is prox-regular at a point $\bar{x} \in S$ if each point close to $\bar{x}$ has a unique nearest point in S. Equivalently, using the terminology of Clarke et al. [5], there exists a scalar $\rho > 0$ such that for all points $x \in S$ close to $\bar{x}$, every unit vector $n \in N_S^P(x)$ can be *realized by a ρ-ball*, meaning

$$S \cap \operatorname{int}(x + \rho(n + B)) = \varnothing$$

(see Poliquin et al. [9, Theorem 1.3]). This property implies Clarke regularity of S at $\bar{x}$ (see Poliquin et al. [9]).

Our aim is to prove the following characterization of the normal cone $N_S(\bar{x})$.

Theorem 4.1 (Normals Without Convexification). *Consider a closed set $S \subset \mathbf{R}^n$ and a point $\bar{x} \in S$. For fixed scalars $\gamma > 4$ and $\alpha \in [0, 1]$, and for each scalar $\epsilon > 0$, define the set N_ϵ to be*

$$\left\{\lambda w\colon\ \lambda \geq 0,\ \ w \in x - P_{S \cap (\bar{x} + \gamma\epsilon B)}(x),\ \ \|x - \bar{x}\| \leq 2\epsilon,\ \ \|w\| \geq \alpha\epsilon\right\} \cup \{0\}.$$

If either $\alpha = 0$ or S is prox-regular at $\bar{x}$, then

$$\lim_{\epsilon \downarrow 0} N_\epsilon = N_S(\bar{x}).$$

PROOF. Without loss of generality we can suppose the point $\bar{x}$ is 0. We first prove the inclusion

$$N_\epsilon \subset \bigcup_{y\in S\cap 4\epsilon B} N_S^P(y). \tag{21}$$

To this end, consider a point $x \in 2\epsilon B$, a point $y \in P_{S\cap\gamma\epsilon B}(x)$, and a scalar $\lambda \geq 0$. If we now define $w = x - y$, then λw is a typical element of the set N_ϵ. (Clearly, 0 belongs to the right-hand side of inclusion (21).)

Since $0 \in S\cap\gamma\epsilon B$, we must have $\|x-y\| \leq \|x\|$, so $\|y\| \leq 2\|x\| \leq 4\epsilon$: that is, $y \in S\cap 4\epsilon B$. For any scalar $\tau \in (0, 1)$, we know

$$P_{S\cap\gamma\epsilon B}(y+\tau w) = \{y\}. \tag{22}$$

We now claim that, providing τ is small, we have

$$P_S(y+\tau w) = \{y\},$$

and, hence, $w \in N_S^P(y)$, proving the desired inclusion. In light of equation (22), it suffices to show that any point $z \notin \gamma\epsilon B$ satisfies

$$\|(y+\tau w) - z\| > \|(y+\tau w) - y\|.$$

To see this, note that

$$\|(y+\tau w) - z\| - \|\tau w\| \geq \|y - z\| - 2\tau\|w\| > (\gamma - 4)\epsilon - 2\tau\|w\|,$$

and the right-hand side is positive for small $\tau > 0$, as desired.

We next prove that, if either $\alpha = 0$ or S is prox-regular at 0, then

$$N_\epsilon \supset \bigcup_{y\in S\cap\epsilon B} N_S^P(y), \tag{23}$$

providing ϵ is sufficiently small. Suppose first that S is prox-regular at 0, so that there exist scalars $\bar{\epsilon}, \rho > 0$ such that for all points $y \in S\cap\bar{\epsilon}B$ and all unit vectors $n \in N_S^P(y)$, we have

$$S \cap \operatorname{int}(y + \rho(n+B)) = \varnothing. \tag{24}$$

We will show that (23) holds for any scalar $\epsilon \in (0, \bar{\epsilon})$ satisfying $(\alpha+1)\epsilon < 2\rho$.

For such an ϵ, define a scalar $\tau = \epsilon(\alpha+1)/2$, so that $0 < \tau < \rho$. Now consider any point $y \in S\cap\epsilon B$ and any normal vector $n \in N_S^P(y)$. We want to show $n \in N_\epsilon$. If $n = 0$, there is nothing to prove, so we can assume $\|n\| = 1$, since both sets $N_S^P(y)$ and N_ϵ are cones. Hence, equation (24) holds, which guarantees that the point $x = y + \tau n$ satisfies $P_S(x) = \{y\}$ and, hence, $P_{S\cap\gamma\epsilon B}(x) = \{y\}$. Clearly, $x \in 2\epsilon B$, and the vector $w = x - y$ has norm $\tau \geq \alpha\epsilon$, as required.

This completes the proof of inclusion (23) in the prox-regular case. In the case $\alpha = 0$, the argument simplifies, since we can choose the scalar $\tau > 0$ in the above paragraph arbitrarily small. To summarize, we now have the inclusions

$$\bigcup_{y\in S\cap\epsilon B} N_S^P(y) \subset N_\epsilon \subset \bigcup_{y\in S\cap 4\epsilon B} N_S^P(y) \tag{25}$$

for all small $\epsilon > 0$.

To prove our desired result, consider an arbitrary sequence of scalars $\epsilon_k \downarrow 0$. First, consider any normal vector $d \in N_S(0)$. By definition, there exist sequences of vectors $y_m \in S$ and $d_m \in N_S^P(y_m)$ (for $m = 1, 2, \ldots$) such that $y_m \to 0$ and $d_m \to d$ as $m \to \infty$. Choose an increasing sequence of natural numbers $m_1 < m_2 < \cdots$ such that $y_{m_k} \in \epsilon_k B$ for each $k = 1, 2, \ldots$. Now, using the left-hand inclusion of (25), we know $d_{m_k} \in N_{\epsilon_k}$ for each k and $d_{m_k} \to d$ as $k \to \infty$. Thus, we have shown $\liminf_{\epsilon\downarrow 0} N_\epsilon \supset N_S(0)$.

On the other hand, suppose the vectors $d_k \in N_{\epsilon_k}$ (for $k = 1, 2, \ldots$) satisfy $d_k \to d$ as $k \to \infty$. Using the right hand inclusion of (25), for each k there exists a point $y_k \in S\cap 4\epsilon_k B$ such that $d_k \in N_S^P(y_k)$. Since $y_k \to 0$ as $k \to \infty$, we deduce $d \in N_S(0)$. We have thus shown $\limsup_k N_k \subset N_S(0)$, which completes the proof. □

5. Normal estimates with convexification. To justify estimates of normal cones involving convexification, like formula (12), we need some simple tools for dealing with Clarke regular sets. The following result is crucial to our approach.

PROPOSITION 5.1 (REGULARITY). *If the set $S \subset \mathbf{R}^n$ is Clarke regular at* 0*, then the tangent and normal cones $T_S(0)$ and $N_S(0)$ are convex and mutually polar, and furthermore*

$$\lim_{\epsilon \downarrow 0} \operatorname{conv}\left(B \cap \frac{1}{\epsilon} S\right) = B \cap T_S(0).$$

PROOF. By Rockafellar and Wets [10, Corollary 6.30], we see the convexity and mutual polarity, and furthermore that S is geometrically derivable at 0. We can therefore deduce

$$\lim_{\epsilon \downarrow 0} B \cap \frac{1}{\epsilon} S = B \cap T_S(0)$$

(see Rockafellar and Wets [10, p. 198]). Being regular, S is locally closed at 0, so using compactness and Rockafellar and Wets [10, Proposition 4.30], we deduce

$$\lim_{\epsilon \downarrow 0} \operatorname{conv}\left(B \cap \frac{1}{\epsilon} S\right) = \operatorname{conv}(B \cap T_S(0)).$$

The right hand side is just $B \cap T_S(0)$, since $T_S(0)$ is convex. □

The next result guarantees that the projection operation in which we are interested does provide good approximations to every normal vector.

PROPOSITION 5.2 (OUTER APPROXIMATION). *Consider a set $S \subset \mathbf{R}^n$ that is Clarke regular at* 0*, and a normal vector $d \in N_S(0)$. Then*

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon}\big(\epsilon d - P_{\operatorname{conv}(S \cap \epsilon B)}(\epsilon d)\big) = d.$$

PROOF. Proposition 5.1 (regularity) shows

$$\lim_{\epsilon \downarrow 0} \operatorname{conv}\left(B \cap \frac{1}{\epsilon} S\right) = B \cap T_S(0).$$

Since $T_S(0)$ and $N_S(0)$ are mutually polar, and $d \in N_S(0)$, we have

$$P_{B \cap T_S(0)}(d) = 0.$$

Applying Rockafellar and Wets [10, Proposition 4.9] now shows

$$\lim_{\epsilon \downarrow 0} P_{\operatorname{conv}(B \cap \epsilon^{-1} S)}(d) = P_{B \cap T_S(0)}(d) = 0.$$

Hence, the limit we seek is

$$d - \lim_{\epsilon \downarrow 0} P_{\operatorname{conv}(B \cap \epsilon^{-1} S)}(d) = d,$$

as required. □

We need the following routine result.

PROPOSITION 5.3 (CONVERGENCE OF PROJECTIONS). *Consider a sequence of nonempty closed convex sets $C_k \to C$ in $\mathbf{R}^n$, and a sequence of vectors $y_k \to y$ in $\mathbf{R}^n$. Then, $P_{C_k}(y_k) \to P_C(y)$.*

PROOF. It is easy to check that $(C_k - y_k) \to (C - y)$. Hence, by Rockafellar and Wets [10, Proposition 4.9], we deduce

$$P_{C_k}(y_k) - y_k = P_{C_k - y_k}(0) \to P_{C-y}(0) = P_C(y) - y.$$

The result now follows. □

Our last tool looks a little more technical.

PROPOSITION 5.4 (INNER APPROXIMATION). *Consider a set $S \subset \mathbf{R}^n$ that is Clarke regular at 0, a sequence of strictly positive scalars $\epsilon_k \downarrow 0$, and a sequence of vectors $x_k \in \epsilon_k B$ (for $k = 1, 2, \ldots$). Suppose the sequence of vectors $w_k = x_k - P_{\mathrm{conv}\,(S \cap \epsilon_k B)}(x_k)$ satisfies*

$$\liminf_k \frac{\|w_k\|}{\epsilon_k} > 0. \tag{26}$$

Then, any cluster point of the sequence $\{\|w_k\|^{-1} w_k\}$ lies in $N_S(0)$.

PROOF. Without loss of generality, we can suppose there exists a unit vector w such that $\|w_k\|^{-1} w_k \to w$. By taking a subsequence, we can suppose there exists a vector $x \in B$ such that $\epsilon_k^{-1} x_k \to x$. For each $k = 1, 2, \ldots$, define a vector

$$s_k = P_{\mathrm{conv}\,(S \cap \epsilon_k B)}(x_k) = \epsilon_k P_{\mathrm{conv}\,(B \cap \epsilon_k^{-1} S)}(\epsilon_k^{-1} x_k).$$

Taking a further subsequence, we can suppose there exists a vector $s \in B$ such that $\epsilon_k^{-1} s_k \to s$. Notice

$$\frac{1}{\epsilon_k} w_k = \frac{1}{\epsilon_k} x_k - \frac{1}{\epsilon_k} s_k \to x - s, \tag{27}$$

so our assumption (26) implies $s \neq x$.

Proposition 5.1 (regularity) shows

$$\mathrm{conv}\,(B \cap \epsilon_k^{-1} S) \to B \cap T_S(0).$$

Using Proposition 5.3 (convergence of projections), we deduce

$$s = P_{B \cap T_S(0)}(x). \tag{28}$$

We first show $\|s\| \neq 1$. Suppose, on the contrary, $\|s\| = 1$. Since we know $s \neq x \in B$, the Cauchy-Schwartz inequality implies $\langle s, x \rangle < 1$. However, since $0 \in B \cap T_S(0)$, we deduce the contradiction

$$0 \leq \langle s, x - s \rangle = \langle s, x \rangle - \|s\|^2 < 1 - 1 = 0.$$

Thus, $\|s\| < 1$.

Equation (28) shows

$$x - s \in N_{B \cap T_S(0)}(s) = N_{T_S(0)}(s) \subset N_S(0),$$

using the mutual polarity of the cones $T_S(0)$ and $N_S(0)$ and the fact that $s \in \mathrm{int}\, B$. Lastly, equation (27) implies

$$w = \lim \frac{1}{\|w_k\|} w_k = \lim \frac{1}{\|\epsilon_k^{-1} w_k\|} \epsilon_k^{-1} w_k = \frac{1}{\|x - s\|}(x - s) \in N_S(0),$$

as desired. □

We are now ready for the main result of this section. Unlike Theorem 4.1 (normals without convexification), while excluding small displacements from consideration, it requires no assumption of prox-regularity.

THEOREM 5.1 (NORMALS WITH CONVEXIFICATION). *Consider a set $S \subset \mathbf{R}^n$ that is Clarke regular at the point $\bar{x} \in S$. Fix a scalar α in the interval $(0, 1)$, and for each scalar $\epsilon > 0$ define the set G_ϵ to be*

$$\{\lambda w\colon\ \lambda \geq 0,\ w = x - P_{\mathrm{conv}\,(S \cap (\bar{x} + \epsilon B))}(x),\ \|x - \bar{x}\| \leq \epsilon,\ \|w\| \geq \alpha\epsilon\} \cup \{0\}.$$

Then,

$$\lim_{\epsilon\downarrow 0} G_\epsilon = N_S(\bar{x}).$$

If, furthermore, S is epi-Lipschitz at $\bar{x}$, then

$$\lim_{\epsilon\downarrow 0} \operatorname{conv} G_\epsilon = N_S(\bar{x}).$$

PROOF. We can assume $\bar{x} = 0$. We first prove

$$N_S(\bar{x}) \subset \liminf_{\epsilon\downarrow 0} G_\epsilon.$$

Clearly, the point 0 belongs to the left-hand side. Since both sides of the inclusion are cones, it therefore suffices to show that for any unit vector $d \in N_S(0)$ there exist vectors $n_\epsilon \in G_\epsilon$ (for all small $\epsilon > 0$) satisfying $\lim_{\epsilon\downarrow 0} n_\epsilon = d$. To see this, define vectors $x_\epsilon = \epsilon d$ and

$$d_\epsilon = x_\epsilon - P_{\operatorname{conv}(S\cap \epsilon B)}(x_\epsilon).$$

Clearly, $\|x_\epsilon\| \le \epsilon$. Proposition 5.2 (outer approximation) shows $\lim_{\epsilon\downarrow 0} \epsilon^{-1} d_\epsilon = d$, so for all small $\epsilon > 0$ we know $\|d_\epsilon\| > \alpha\epsilon$, and hence the vector $n_\epsilon = \|d_\epsilon\|^{-1} d_\epsilon$ lies in the set G_ϵ. However, now clearly $\lim_{\epsilon\downarrow 0} n_\epsilon = d$, as required.

We next turn to the opposite inclusion:

$$\limsup_{\epsilon\downarrow 0} G_\epsilon \subset N_S(\bar{x}).$$

Consider any sequences of scalars $0 < \epsilon_k \downarrow 0$ and $\lambda_k \ge 0$, and any sequence of vectors $x_k \in \epsilon_k B$ (for $k = 1, 2, \ldots$) such that the sequence of vectors

$$w_k = x_k - P_{\operatorname{conv}(S\cap \epsilon_k B)}(x_k)$$

satisfy $\|w_k\| \ge \alpha\epsilon_k$ for all k and $\lambda_k w_k \to w$. We want to show $w \in N_S(0)$. If $w = 0$, there is nothing to prove, so suppose $w \ne 0$, and hence $\|w_k\|^{-1} w_k \to \|w\|^{-1} w$. Then, $\liminf_k \epsilon_k^{-1} \|w_k\| \ge \alpha > 0$, so Proposition 5.4 (inner approximation) shows $\|w\|^{-1} w$ lies in $N_S(0)$, and hence so does w.

We have now proved the first statement of the theorem. The final statement now follows (see Rockafellar and Wets [10, Proposition 4.30]), since the normal cone $N_S(\bar{x})$ is pointed as a consequence of the epi-Lipschitz assumption. □

We end with an example to illustrate the necessity of assuming $\alpha > 0$. Consider the closed set

$$S = \{(\gamma, \nu) \in \mathbf{R}^2 \colon \nu = \pm\gamma^2\}.$$

The proximal normal cone at each point in S is easily checked to be

$$N_S^P(\gamma, \gamma^2) = \mathbf{R}(-2\gamma, 1)$$

$$N_S^P(\gamma, -\gamma^2) = \mathbf{R}(2\gamma, 1),$$

for all $\gamma \in \mathbf{R}$, and hence S is everywhere Clarke regular with, in particular,

$$N_S(0, 0) = \mathbf{R}(0, 1).$$

For any scalar $\epsilon > 0$, the set $S_\epsilon = \operatorname{conv}(S \cap \epsilon B)$ is the rectangle with vertices $(\pm\alpha, \pm\alpha^2)$, where α is the positive root of the equation $\alpha^4 + \alpha^2 = \epsilon^2$. It is easy to check that the set

$$\{\lambda(x - P_{S_\epsilon}(x)) \colon \lambda \ge 0, \|x\| \le \epsilon\}$$

is just the union of the two coordinate axes $\{(\gamma, \nu) \colon \gamma\nu = 0\}$, independent of ϵ. This set is strictly larger than $N_S(0, 0)$, so Theorem 5.1 (normals via projections) can fail if we allow $\alpha = 0$.

6. Numerical tests. In this section we examine some simple numerical tests on normal cone estimation. In §§4 and 5 we saw two methods of estimating the normal cone to a set without using calculus. To put these two methods into the same framework, we consider the following algorithm. We emphasize that this procedure is not meant as a direct optimization tool, but rather as an experiment to study how much random local data can reveal about the normal cone.

Algorithm: Random normal generation (RNG). Given a set S described by an oracle, a feasible point $\bar{x}$, a search radius $\epsilon > 0$, a minimum-projection parameter $\alpha \in [0, 1)$, a maximum-length parameter $\gamma > 0$, and a sample size N, perform the following:

I. Selection. Pick a set of N random points

$$A = \{\tilde{x}_i\colon\ i = 1, 2, \ldots, n\},$$

all within distance ϵ of $\bar{x}$.

II. Organization. Using the oracle, organize these points into two sets:

$$P_{\text{in}} = (A \cap S) \cup \{\bar{x}\}$$

$$P_{\text{out}} = \{x \in A \cap S^c\colon\ \gamma\|x - \bar{x}\| \leq 2\epsilon\}.$$

III. Estimation. Use P_{in} to calculate an estimate $\widetilde{S}$ of S.

IV. Projection. Project each point in P_{out} onto $\widetilde{S}$ to approximate the normal cone:

$$\widetilde{N} = \mathbf{R}_+ \operatorname{conv}\left\{x - P_{\widetilde{S}}(x)\colon\ x \in P_{\text{out}},\ \|x - P_{\widetilde{S}}(x)\| > \alpha\epsilon\right\}.$$

In this framework the two approaches differ only in how they perform Step III (Estimation). The first approach (§4) estimates the set S by simply using the collection of points found to be in the set, P_{in}. This makes the calculation in Step IV (Projection) extremely simple. The second approach (§5) uses the more complicated approximation $\widetilde{S} = \operatorname{conv} P_{\text{in}}$. In this case the calculation of the projection mapping becomes a quadratic program. Henceforth we shall call these two methods the *convex hull-free* (**CHF**) and the *convex hull* (**CH**) method.

Examining Sections 4 and 5 it is clear that the **CH** method has several theoretical advantages. The **CHF** method (theoretically) requires the maximum-length parameter, γ, to be greater than 4 (see Theorem 4.1). It further (theoretically) requires either the set to be prox-regular, or the minimum-projection parameter, α, to be 0. Setting the minimum-projection parameter to 0 is undesirable as it reduces the robustness of the algorithm.

In the **CH** method the maximum length parameter can be set to 1, effectively removing it from the calculations (see Theorem 5.1). Also, prox-regularity is no longer required. However, these improvements come at the price of complicating the calculation in Step IV (Projection) of the algorithm.

In this section we examine two simple numerical tests comparing the **CHF** and **CH** method of normal cone approximation. Since Theorems 4.1 and 5.1 suggest that the accuracy of the approximated normal cone is effected by the search radius, ϵ, and the number of points selected, N, we also take this opportunity to briefly examine the effect of adjusting these parameters on the resulting approximation.

Our first test considers the effect of increasing the number of points selected, N, on the accuracy of the approximate normal cone. To do this we consider the positive orthant, $\mathbf{R}^n_+$, at the origin. We attempt to reconstruct the normal cone by randomly selecting 10, 100, and 1,000 points, then applying the **CH** and **CHF** method. We calculate the error in the approximate normal cone via the following formula:

$$\max\left\{\begin{array}{l}\max\{\arcsin(\|w - P_{N_S(\bar{x})}(w)\|)\colon\ w \in \widetilde{N},\ \|w\| = 1\}\\ \max\{\arcsin(\|w - P_{\widetilde{N}}(w)\|)\colon\ w \in N_S(\bar{x}),\ \|w\| = 1\}\end{array}\right\}, \tag{29}$$

TABLE 1. Mean and standard deviations for RNG on positive orthant.

Problem	Dimension and approach	No. of points	Mean error	Standard deviation
Positive orthant	2, **CH**	10	0.8133	0.3671
		100	0.4208	0.2635
		1000	0.2059	0.1177
$\bar{x} = 0$	2, **CHF**	10	0.9529	0.3644
		100	0.8350	0.2296
		1000	0.4787	0.0880
	5, **CH**	10	1.2852	0.1960
		100	1.1166	0.1613
		1000	0.9886	0.1056
	5, **CHF**	10	1.3081	0.1797
		100	1.3916	0.1231
		1000	1.3713	0.0879

(where $N_S(\bar{x})$ is the correct normal cone, and $\widetilde{N}$ is the approximate normal cone). Formula (29) measures the maximum angle between correct and approximate normal cones. In this case, the correct normal cone is the negative orthant ($N_S(\bar{x}) = -\mathbf{R}^n_+$).

In order to counter the randomness of the algorithm, we run each test 100 times. Table 1 displays the mean and standard deviation of the resulting errors, while Figure 1 provides a histogram of the error frequencies.

The results for this series of tests shows several interesting trends. First, and least surprising, as the dimension of the problem increases, the accuracy of the approximate normal cone decreases. Second, when convexification is applied, increasing the number of points

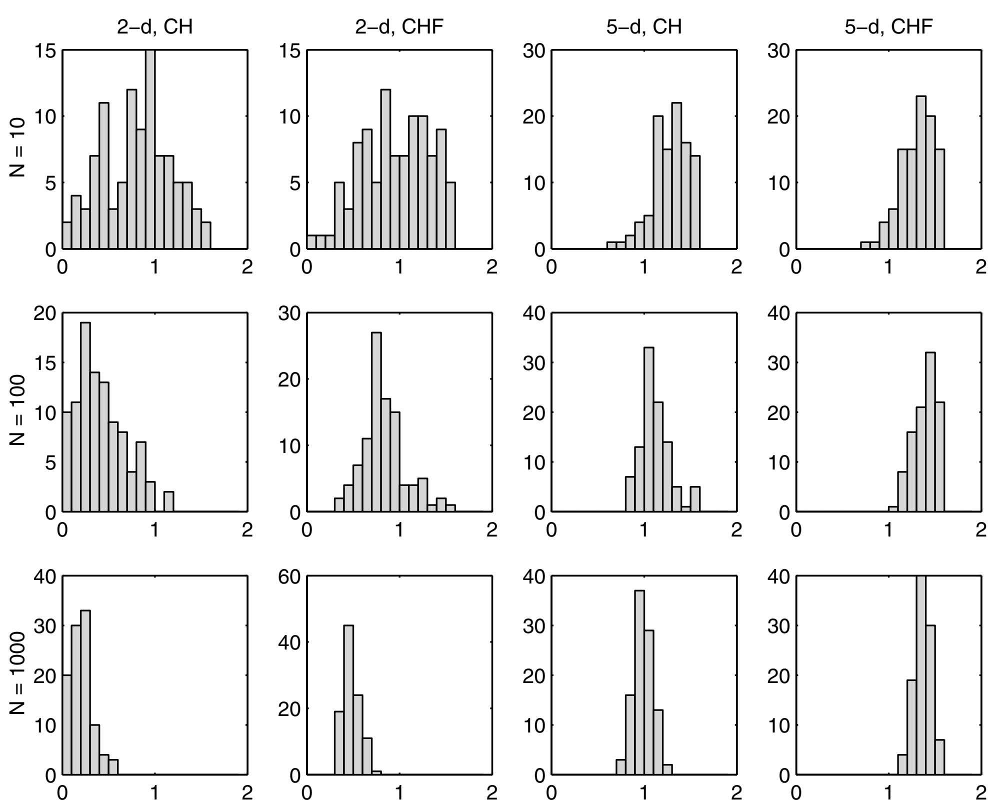

FIGURE 1. Effects of sample size and dimension.

TABLE 2. Mean and standard deviations for RNG on complement of ball.

Problem	Dimension and approach	Search radius	Mean error	Standard deviation
Complement of ball	2, **CH**	1.25	0.0850	0.0724
		0.50	0.2189	0.2138
		0.05	0.3827	0.2774
$\bar{x} = (1, 0, 0, \ldots, 0)$	2, **CHF**	1.25	1.3334	0.2155
		0.50	0.9393	0.2506
		0.05	0.8367	0.2387
	5, **CH**	1.25	0.7629	0.1732
		0.50	0.9988	0.1274
		0.05	1.1602	0.1215
	5, **CHF**	1.25	1.5044	0.1100
		0.50	1.4444	0.0993
		0.05	1.4301	0.0772

increases the accuracy of the normal cone produced. When no convexification is applied, increasing the number of points selected seems to have somewhat less effect on the accuracy of the approximate normal cone.

Our second test considers the effect of reducing the search radius, ϵ, on the accuracy of the approximate normal cone. In this case we consider the complement of the open ball $S = \{x\colon \|x\| \geq 1\}$ at the point $\bar{x} = (1, 0, 0, \ldots, 0)$. As before we consider the test in two and five dimensions. In this case, however, instead of altering the number of points selected, we always use 100 randomly selected points and repeat the test for various search radii. As before, we run each test 100 times: Table 2 and Figure 2 display the results.

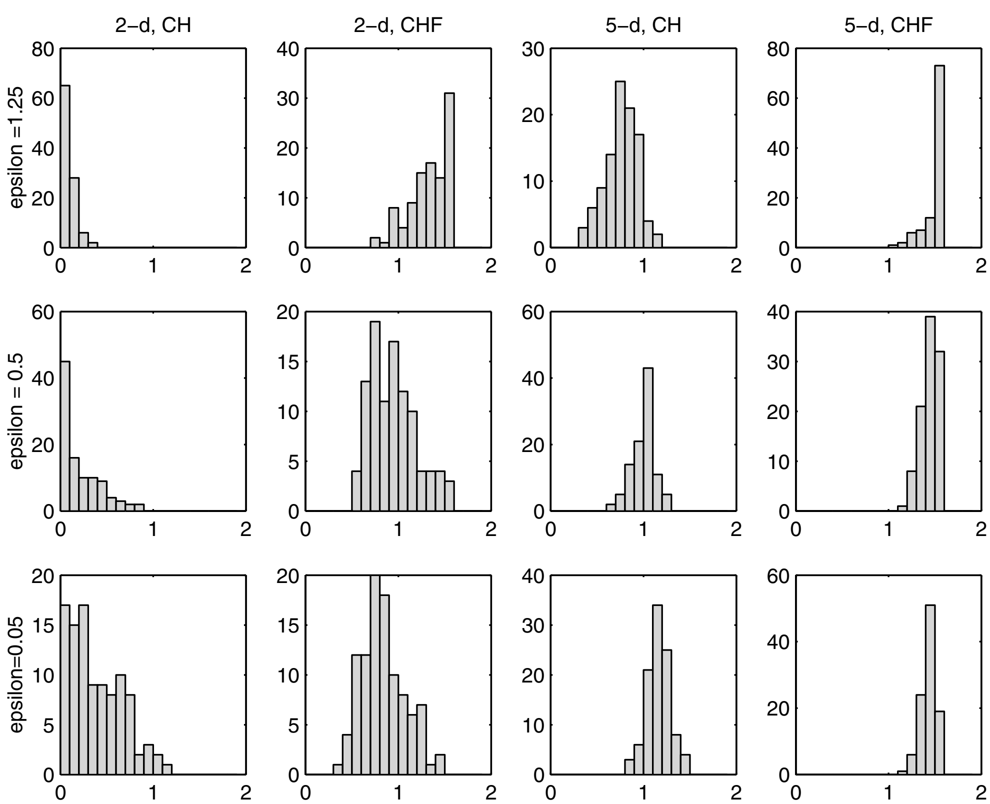

FIGURE 2. Effects of sampling radius and dimension.

Examining Table 2 reveals several results. We once again see that taking the convex hull creates a better estimate for the normal cone than the **CHF** method. This suggests that the **CH** method is making better use of the data given by the oracle. In cases where checking feasibility is difficult, this may be important.

The data also show that when the search radius is large, the normal cone produced by the **CHF** method is less accurate. This is expected, as the algorithm picks up normal vectors for points within the radius. For a large radius, the algorithm creates normal vectors to points further from the central point, causing a larger approximate normal cone. As the radius decreases, one expects the approximate normal cone to become more accurate.

Surprisingly, this trend does not continue for the **CH** method. In fact, the errors from testing with the convexification step are actually smaller when the search radius is large. A moment's thought may explain this phenomenon. The convexification step causes the approximation of the set to become polyhedral, including a large face enclosing many points outside the set. The normal vectors created will only include those produced from points outside this face. These will be projected onto this face, causing (by good fortune) a high accuracy in the approximate normal cone. When the convexification step is skipped, the approximate normal cone does not have this advantage and is considerably less accurate.

As before, the second series of tests suggests that as problem dimensions increase the difficulty in approximating the normal cone increases. The inaccuracy of the results in just five dimensions already suggests that the **CH** and **CHF** methods are somewhat impractical. It is worth remembering however that the accuracy is measured solely by the worst-case scenario. Therefore, a single outlying vector can cause a large error even if the rest of the approximate normal cone is accurate. Clearly, more refinement is needed before these approximation ideas become a practical optimization tool.

References

[1] Burke, J. V., A. S. Lewis, M. L. Overton. 2002. Approximating subdifferentials by random sampling of gradients. *Math. Oper. Res.* **27** 567–584.

[2] Burke, J. V., A. S. Lewis, M. L. Overton. 2005. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. Optim.* **15** 751–779.

[3] Clarke, F. H. 1973. Necessary conditions for nonsmooth problems in optimal control and the calculus of variations. Ph.D. thesis, University of Washington, Seattle, WA.

[4] Clarke, F. H., Yu. S. Ledyaev, R. J. Stern, P. R. Wolenski. 1998. *Nonsmooth Analysis and Control Theory*. Springer-Verlag, New York.

[5] Clarke, F. H., R. J. Stern, P. R. Wolenski. 1995. Proximal smoothness and the lower-C^2 property. *J. Convex Anal.* **2** 117–144.

[6] Ioffe, A. D. 1981. Sous-différentielles approchées de fonctions numériques. *Comptes Rendus Académie Sciences Paris* **292** 675–678.

[7] Kruger, A. Y., B. S. Mordukhovich. 1980. Extremal points and the Euler equation in nonsmooth analysis. *Dokl. Akad. Nauk BSSR* **24** 684–687.

[8] Mordukhovich, B. S. 1976. Maximum principle in the problem of time optimal response with nonsmooth constraints. *J. Appl. Math. Mech.* **40** 960–969.

[9] Poliquin, R. A., R. T. Rockafellar, L. Thibault. 2000. Local differentiability of distance functions. *Trans. Amer. Math. Soc.* **352** 5231–5249.

[10] Rockafellar, R. T., R. J.-B. Wets. 1998. *Variational Analysis.* Springer, Berlin, Germany.