Exploiting Known Structures to Approximate Normal Cones

C. Davis* W. Hare†

February 16, 2012

Abstract

The normal cone to a constraint set plays a key role in optimization theory, algorithms, and applications. We consider the question of how to approximate the normal cone to a set under the assumption that the set is provided through an oracle function or collection of oracle functions, but contains some exploitable structure. We provide a new simplex gradient based approximation technique that works for sets of the form

\[ S = \{x | g_i(x) \leq 0, i = 1, \ldots, N\}, \]

where each \( g_i : \mathbb{R}^n \to \mathbb{R} \) is unknown and provided by an oracle. We further present novel results showing that, under a non-degeneracy condition, approximating normal cones to intersections of sets is possible by taking sums of approximations. Finally, we provide numerical results that exemplify the accuracy of the simplex gradient approximation when it is applicable, and the fail of this technique when linear independence constraint qualification is not met.

Keywords: Nonsmooth optimization, normal cone, constraint set, numerical analysis, discrete approximate gradient, oracle

AMS Classification Primary: 49M25, 65K10; Secondary: 49J52, 58C06, 90C46

1 Introduction

When dealing with smooth functions, gradients play a prominent role in the field of optimization. For nonsmooth functions this role is subsumed by “subgradients” and the “subdifferential map”. When dealing with sets, the analogue of these objects are the cones made up of normal vectors to the set, appropriately dubbed normal cones. Before proceeding further, it is prudent to formally define the normal cone now.

**Definition 1.1 (normal cone)** Let \( S \subseteq \mathbb{R}^n \). Then the proximal normal cone to \( S \) at a point \( \bar{x} \in S \) is given by

\[ N^P_S(\bar{x}) = \{\lambda(y - \bar{x}) | \lambda \geq 0, \bar{x} \in P_S(y)\} \]

where \( P_S(y) \) denotes the projection of \( y \) onto \( S \): \( P_S(y) = \{x \in S : |x - y| = \inf_{x' \in S} |x' - y|\} \).

The (limiting) normal cone to \( S \) at a point \( x \in S \) is the cone

\[ N_S(x) = \{w \mid \text{there exists } x_k \in S, w_k \in N^P_S(x_k) \text{ with } x_k \to x, w_k \to w\}. \]

Simply put, the normal cone consists of all vectors that point orthogonally away from the set.

In considering the minimization of a smooth function \( f \) over a constraint set \( S \), the normal cone provides the “first order necessity criterion”,

\[ \bar{x} \in \arg\min \{f(x) : x \in S\} \Rightarrow -\nabla f(\bar{x}) \in N_S(\bar{x}), \]

---


†Mathematics. University of British Columbia, Okanagan Campus. warren.hare@ubc.ca. Research supported in part by NSERC DG #355571-2008.
which can further be used to derive the celebrated KKT conditions (see [DD12, Thm 3.7] for example). When \( f \) and \( S \) are both convex, equation (1) statement becomes both necessary and sufficient to detect a global minimizer.

When working with a nonsmooth function \( f \), the normal cone can be used to define the subdifferential map at a point \( \bar{x} \):

\[
\bar{v} \in \partial f(\bar{x}) \iff (\bar{v}, -1) \in N_{\text{epi}(f)}(\bar{x}, f(\bar{x}))
\]

(see for example [Cla90, eq (10), p. 13] or [RW98, Thm 8.9]). In nonsmooth optimization, the subdifferential map subsumes the role of a gradient map and is vital to both theory and algorithm design. Classical research on approximating gradients maps without analytic knowledge of the function is well established [Kel99]. More recently, research has begun to examine both how to approximate subdifferential maps and how to employ them in a derivative-free optimization setting [BKS08] [CSV09] [Kiw10] [HM11] [HN12]. Yet very little research has explored how to numerically approximate normal cones.

In [HL05], Hare and Lewis presented a first study on how to approximate the normal cone to a set without an analytic description of the set. In particular, they considered the situation where the set was given by an oracle that could be consulted to determine if a given point was inside or outside of the set. Such a situation may arise when feasibility of a constraint is determined through a computer simulation or other such black-box.

However, in practice constraint sets are seldom so unstructured. Constraint sets may take the form of a finitely constrained set:

\[
S = \{ x : g_i(x) \leq 0, \text{ for } i = 1, 2, \ldots m \},
\]

where each \( g_i \) is smooth function, or constraint sets may take the form of an intersection of various sets:

\[
S = S_1 \cap S_2 \cap \ldots \cap S_n.
\]

In the latter case, it may be that some constraint sets are very well understood (and a true normal cone is obtainable), while others are generated through oracle functions.

In this work we consider these scenarios, and develop several methods to exploit structure within a constraint set that is (at least partially) given by oracle functions. We begin by considering finitely constrained sets where each \( g_i \) is given by a oracle, that is,

\[
S = \{ x : g_i(x) \leq 0, \text{ for } i = 1, 2, \ldots m \},
\]

and given a point \( x \), we can consult the oracle to determine a function value \( g_i(x) \), but gradients and other analytic information regarding \( g_i \) are unavailable. Although the techniques from [HL05] are applicable to this setting, it is clear that such a set contains a powerful structure that might be exploited to generate better approximations. In particular, we provide a novel simplex gradient based technique to approximate normal cones in this setting. We provide both theoretical convergence results (in Section 2) and numerical tests (in Section 4) that demonstrate both the strengths and weaknesses of this new technique.

We also consider the situation where the constraint set is a finite intersection of various sets. In Section 3 we show that, under a non-degeneracy condition, normal cones for the intersection can be approximated by way of approximate normal cones to each individual set. This will be particularly useful when one (or more) of the sets are analytically available, while others are not.

Before continuing we provide a brief reminder on the definitions and theory behind the simplex gradient. Further information and applications of the simplex gradient can be found in [Kel99] and [CSV09].

### 1.1 Simplex Gradients

Given a set \( S \), the convex hull of \( S \) is the the smallest convex set containing \( S \) and denoted \( \text{conv}(S) \). The conic hull of \( S \) is the the smallest conic set containing \( S \) and denoted \( \text{cone}(S) \). Given a set \( Y \) that consists of
a finite number of points $Y = \{y_j\}_{j=0}^n$, the convex hull, conic hull, and convex conic hull, can be generated via the following formulae.

$$\text{conv}(Y) = \{x : x = \sum_{i=0}^n \lambda_i y_i, \quad \sum_{i=0}^n \lambda_i = 1, \quad \lambda_i \geq 0, \quad y_i \in Y\},$$

$$\text{cone}(Y) = \{x : x = \lambda_i y_j, \quad \text{for some} \quad \lambda_i \geq 0, \quad y_j \in Y\},$$

$$\text{conv}(\text{cone}(Y)) = \{x : x = \sum_{i=0}^n \lambda_i y_j, \quad \lambda_i \geq 0\}.$$

To utilize the simplex gradient, we must first define a simplex.

**Definition 1.2 (simplex)** A simplex in $\mathbb{R}^n$ is the convex hull of $n+1$ points, $Y = \{y_j\}_{j=0}^n$, whose interior is nonempty. We say $\text{conv}(Y)$ is the simplex generated by $Y$ and $Y$ forms the simplex $\text{conv}(Y)$.

The simplex gradient represents the gradient of the linear function generated by interpolation of $f$ over the simplex generating points $Y$. Formally we have.

**Definition 1.3 (simplex gradient)** Let $Y = \{y_j\}_{j=0}^n$ form a simplex in $\mathbb{R}^n$. The simplex gradient $\nabla_s f(Y)$ is given by

$$\nabla_s f(Y) = V^{-T} \delta(f : Y)$$

where $V$ is the matrix of simplex directions

$$V = [y_0 - y_1, \ldots, y_0 - y_n]$$

and $\delta(f : Y)$ is the vector of objective function differences

$$\delta(f : Y) = (f(y_1) - f(y_0), f(y_2) - f(y_0), \ldots, f(y_n) - f(y_0))^T.$$

We next define the simplex radius and condition number, both of which will be used extensively in examining the accuracy of our normal cone approximation.

**Definition 1.4 (simplex radius and condition number)** Let $Y = \{y_j\}_{j=0}^n$ form a simplex in $\mathbb{R}^n$. The simplex radius of $Y$ is defined by

$$\Delta Y = \max_{1 \leq i \leq n} |y_0 - y_i|.$$

The condition number of $Y$ is defined by

$$\kappa(Y) = \|V\| \|V^{-1}\|$$

where $V$ is the matrix of simplex directions (equation (2)).

Using these two definitions, we give our first lemma.

**Lemma 1.5** Let $Y = \{y_0, y_1, \ldots, y_n\}$ form a simplex. Define

$$Y_\varepsilon = \{y_0, \varepsilon(y_1 - y_0) + y_0, \ldots, \varepsilon(y_n - y_0) + y_0\}.$$  

Then $Y_\varepsilon$ forms a simplex with $\Delta Y_\varepsilon = \varepsilon \Delta Y$ and $\kappa(Y_\varepsilon) = \kappa(Y)$.

**Proof:** The proof follows directly from definitions 1.6 and 1.7. \qed

In order to approximate the normal cone via the simplex gradient, we shall require the knowledge that the simplex gradient provides a good approximation of the true gradient.

**Proposition 1.6 (simplex gradient error bounds)** Let $Y = \{y_0, \ldots, y_n\}$ form a simplex in $\mathbb{R}^n$. Suppose that $f \in C^1$ and $\nabla f$ is Lipschitz continuous on $B_{\Delta Y}(y_0)$ with Lipschitz constant $2K_f$. Then

$$|\nabla f(x) - \nabla_s f(Y)| \leq \sqrt{n} K_f \kappa(Y) \Delta Y \quad \text{for all} \quad x \in B_{\Delta Y}(y_0).$$

**Proof:** [Kel99, Lem 6.2.1] \qed
2 Finitely Constrained Sets

We now turn our attention to approximating normal cones to finitely constrained sets that are provided by black-box functions. Because we will be working with finitely constrained sets, to approximate the normal cone with the simplex gradient we shall impose that the Linear Independence Constraint Qualification (LICQ) be met. We first define the active set.

**Definition 2.1 (active set)** Let

\[ S = \{ x | g_i(x) \leq 0, i = 1, 2, \ldots, N \} \]

where \( g_i \in C^1 \) for all \( i = 1, 2, \ldots, N \). Let \( \bar{x} \in S \). Then the active set of \( S \) at \( \bar{x} \) is

\[ A(\bar{x}) = \{ i | g_i(\bar{x}) = 0, i = \{1, 2, \ldots, N\} \} \].

**Definition 2.2 (LICQ)** Let \( S = \{ x | g_i(x) \leq 0, i = 1, 2, \ldots, N \} \). We say that \( S \) satisfies the linear independence constraint qualification at \( \bar{x} \) if the set of active gradients \( \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \) is a linearly independent set.

Under LICQ it is possible to write the normal cone to a finitely constrained set as the convex conic hull of the active gradients.

**Theorem 2.3 (LICQ and the normal cone)** Let \( S = \{ x | g_i(x) \leq 0, i = 1, 2, \ldots, N \} \). Suppose \( S \) satisfies LICQ at the point \( \bar{x} \in S \). Then

\[ N_S(\bar{x}) = \text{conv} \{ \text{cone} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \}. \tag{4} \]

**Proof:** [RW98, Thm 6.14]

It is this theorem that will allow us to use simplex gradients to approximate the normal cone to a finitely constrained set. Essentially we shall replace the exact gradients in equation (4) with simplex gradients and show that the conic approximation remains accurate. However, before full details can be explained, several lemmas will be required.

**Lemma 2.4** Let \( S = \{ x | g_i(x) \leq 0, i = 1, \ldots, N \} \), where \( g_i : \mathbb{R}^n \to \mathbb{R} \). Suppose \( \bar{x} \in S \). Let \( A(\bar{x}) \) be the active set of \( S \) at \( \bar{x} \). Let \( Y = \{ y_0, \ldots, y_n \} \) form a simplex with \( y_0 = \bar{x} \). For \( \varepsilon > 0 \), let \( Y_\varepsilon \) be the simplex defined in Lemma 1.5 (equation (3)). Define the following sets

\[ G_\varepsilon = \{ \nabla g_i(Y_\varepsilon) \}_{i \in A(\bar{x})} \quad \text{and} \quad H = \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})}. \]

Then

\[ \lim_{\varepsilon \downarrow 0} G_\varepsilon = H. \]

**Proof:** Note that \( g_i \) smooth implies that \( \nabla f \) is Lipschitz continuous over any compact set. Combining Proposition 1.6 and Lemma 1.5 we see that for each \( i \)

\[ |\nabla g_i(Y_\varepsilon) - \nabla g_i(\bar{x})| \leq K \kappa(Y_\varepsilon) \Delta Y_\varepsilon = \varepsilon (K \kappa(Y) \Delta Y). \]

From here the statement follows easily by letting \( \varepsilon \downarrow 0 \). \( \square \)

The preceding result gives us many of the tools needed to prove our normal cone approximation. However, before we can give the approximation of the normal cone, we must show that for a set \( S \) with LICQ holding at a point \( \bar{x} \in S \), that the distance between the convex cone of the simplex gradient and the normal cone is bounded. This proof will rely on \( \text{dist}(0, \text{conv} (\nabla g_i(Y_\varepsilon))) \) and \( \text{dist}(0, \text{conv} (\nabla g_i(\bar{x}))) \); specifically the proof will rely on \( \text{dist}(0, \text{conv} (\nabla g_i(Y_\varepsilon))) \) converging to \( \text{dist}(0, \text{conv} (\nabla g_i(\bar{x}))) \), which is shown with the following corollary.
Corollary 2.5 Let \( S = \{ x | g_i(x) \leq 0, i = 1, \ldots, N \} \), where \( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \). Suppose \( \bar{x} \in S \). Let \( A(\bar{x}) \) be the active set of \( S \) at \( \bar{x} \). Let \( Y = \{ y_0, \ldots, y_n \} \) form a simplex with \( y_0 = \bar{x} \). For \( \varepsilon > 0 \), let \( Y_\varepsilon \) be the simplex defined in Lemma 1.5 (equation (3)). Then

\[
\lim_{\varepsilon \downarrow 0} (\text{dist}(0, \text{conv} \{ \nabla g_i(Y_\varepsilon) \}_{i \in A(\bar{x})})) = \text{dist}(0, \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})})
\]

Proof: Using \( G_\varepsilon \) and \( H \) as in Lemma 2.4, we have that \( \lim_{\varepsilon \downarrow 0} G_\varepsilon = H \). Therefore by [RW98, Prop 4.30],

\[
\lim_{\varepsilon \downarrow 0} (\text{conv} \{ G_\varepsilon \}) = \text{conv} \{ H \}
\]

for every \( x \in \mathbb{R}^n \). In particular, we have

\[
\lim_{\varepsilon \downarrow 0} (\text{dist} (0, \text{conv} \{ G_\varepsilon \})) = \text{dist} (0, \text{conv} \{ H \})
\]

that is,

\[
\lim_{\varepsilon \downarrow 0} (\text{dist} (0, \text{conv} \{ \nabla g_i(Y_\varepsilon) \}_{i \in A(\bar{x})})) = \text{dist} (0, \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})}).
\]

The proof also relies on \( \text{dist}(0, \text{conv} \{ \nabla g_i(\bar{x}) \}) \neq 0 \), which is shown by the following lemma.

Lemma 2.6 Let \( S = \{ x | g_i(x) \leq 0, i = 1, \ldots, N \} \), where \( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \). Suppose \( \bar{x} \in S \). If LICQ holds at \( \bar{x} \), then \( 0 \notin \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \).

Proof: Let \( A(\bar{x}) \) be the active set of \( S \) at \( \bar{x} \). For eventual contradiction, suppose that LICQ holds at \( \bar{x} \) and that \( 0 \in \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \). As LICQ holds, \( \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \) is a linearly independent set. This implies that

\[
\sum_{i \in A(\bar{x})} \theta_i \nabla g_i(\bar{x}) = 0 \iff \theta_i = 0 \quad \text{for all } i.
\]

As \( 0 \in \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \), we have that \( \sum_{i \in A(\bar{x})} \theta_i \nabla g_i(\bar{x}) = 0 \) for some \( 0 \leq \theta_i \leq 1 \) where \( \sum_{i \in A(\bar{x})} \theta_i = 1 \). This contradicts LICQ, therefore, \( 0 \notin \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \). □

With these two results in place, we may now show that the distance from the convex cone of simplex gradients and the actual normal cone is bounded.

Theorem 2.7 (simplex gradient bound) Let \( S = \{ x | g_i(x) \leq 0, i = 1, \ldots, N \} \) where \( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \). Suppose \( \bar{x} \in S \) and LICQ holds at \( \bar{x} \). Let \( A(\bar{x}) \) be the active set of \( S \) at \( \bar{x} \). Let \( Y = \{ y_0, y_1, \ldots, y_n \} \) form a simplex with \( y_0 = \bar{x} \). For \( \varepsilon > 0 \) let \( Y_\varepsilon \) be the simplex defined in Lemma 1.5 (equation (3)). Define

\[
\tilde{N}_\varepsilon = \text{conv} \{ \text{cone} \{ \nabla g_i(Y) \}_{i \in A(\bar{x})} \}
\]

(5)

If \( \varepsilon \) is sufficiently small, then the following are true.

1. There exists \( K_1 > 0 \) dependent on the local Lipschitz constants of \( \nabla g_i \), the condition number \( \kappa(Y) \), and the distance \( \text{dist}(0, \text{conv} \{ \nabla g_i(\bar{x}) \}) \) such that

\[
\text{dist}(\hat{n}, N_\varepsilon ) \leq K_1 \Delta Y_\varepsilon \quad \text{for all } \hat{n} \in \tilde{N}_\varepsilon \text{ with } |\hat{n}| = 1.
\]

2. There exists \( K_2 > 0 \) dependent on the local Lipschitz constants of \( \nabla g_i \), the condition number \( \kappa(Y) \), and the distance \( \text{dist}(0, \text{conv} \{ \nabla g_i(\bar{x}) \}) \) such that

\[
\text{dist}(\hat{n}, \tilde{N}_\varepsilon ) \leq K_2 \Delta Y_\varepsilon \quad \text{for all } \hat{n} \in N_\varepsilon \text{ with } |\hat{n}| = 1.
\]
Proof: Let $D = \text{dist}(0, \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})})$. Note that, since LICQ holds, Lemma 2.6 implies that $D > 0$. Consider $F(\varepsilon) = \text{dist}(0, \text{conv} \{ \nabla s g_i(Y_\varepsilon) \}_{i \in A(\bar{x})})$. By Corollary 2.5 we have that $\lim_{\varepsilon \to 0} F(\varepsilon) = D$. Therefore, there exists $\bar{\varepsilon} > 0$ such that

$$|F(\varepsilon) - D| < \frac{D}{2} \quad \text{for all} \quad 0 < \varepsilon < \bar{\varepsilon}.$$ 

This implies that $F(\varepsilon) > \frac{D}{2}$ for all $0 < \varepsilon < \bar{\varepsilon}$. We will henceforth assume that $\varepsilon$ is sufficiently small that

$$\text{dist}(0, \text{conv} \{ \nabla s g_i(Y_\varepsilon) \}_{i \in A(\bar{x})}) > \frac{1}{2} \text{dist}(0, \text{conv} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})}). \quad (6)$$

As $\hat{N}_\varepsilon = \text{conv} \{ \text{cone} \{ \nabla s g_i(Y_\varepsilon) \}_{i \in A(\bar{x})} \}$, we know that

$$\hat{N}_\varepsilon = \left\{ n \mid n = \alpha \sum_{i \in A(\bar{x})} (\theta_i \nabla s g_i(Y_\varepsilon)), \sum_{i \in A(\bar{x})} \theta_i = 1, 0 \leq \theta_i \leq 1, \alpha \geq 0 \right\}.$$ 

As LICQ holds at $\bar{x}$, we have that $N_S(\bar{x}) = \text{conv} \{ \text{cone} \{ \nabla g_i(\bar{x}) \}_{i \in A(\bar{x})} \}$, i.e.,

$$N_S(\bar{x}) = \left\{ n \mid n = \beta \sum_{i \in A(\bar{x})} (\lambda_i \nabla g_i(\bar{x})), \sum_{i \in A(\bar{x})} \lambda_i = 1, 0 \leq \lambda_i \leq 1, \beta \geq 0 \right\}.$$ 

Let $\hat{n} \in \hat{N}_\varepsilon$, so $\hat{n} = \alpha \sum_{i \in A(\bar{x})} \theta_i \nabla s g_i(Y_\varepsilon)$ with $|\hat{n}| = 1$, let

$$n = \alpha \sum_{i \in A(\bar{x})} \theta_i \nabla g_i(\bar{x}) \quad (7)$$

where $\alpha$ and $\theta_i$ are defined as they were for $\hat{n}$. Clearly $n \in N_S(\bar{x})$ as $\alpha \geq 0$, $0 \leq \theta_i \leq 1$, and $\sum_{i \in A(\bar{x})} \theta_i = 1$.

We now examine $|\hat{n} - n|$. As $\alpha \geq 0$ and $\theta_i \geq 0$, we have the following,

$$|\hat{n} - n| = |\alpha \sum_{i \in A(\bar{x})} (\theta_i \nabla s g_i(Y_\varepsilon)) - \alpha \sum_{i \in A(\bar{x})} (\theta_i \nabla g_i(\bar{x}))|$$

$$= |\alpha \sum_{i \in A(\bar{x})} (\theta_i \nabla s g_i(Y_\varepsilon)) - \sum_{i \in A(\bar{x})} (\theta_i \nabla g_i(\bar{x}))|$$

$$= |\alpha \sum_{i \in A(\bar{x})} \theta_i (\nabla s g_i(Y_\varepsilon) - \nabla g_i(\bar{x}))|$$

$$\leq \alpha \sum_{i \in A(\bar{x})} \theta_i |\nabla s g_i(Y_\varepsilon) - \nabla g_i(\bar{x})|$$

where the last step is an application of the triangle inequality. As the $g_i$ are smooth, the Lipschitz assumption holds. Thus applying Proposition 1.6, while noting that $\sum_{i \in A(\bar{x})} \theta_i = 1$, we get the following,

$$|\hat{n} - n| \leq \alpha \sum_{i \in A(\bar{x})} \theta_i |\nabla s g_i(Y_\varepsilon) - \nabla g_i(\bar{x})|$$

$$\leq \alpha \sum_{i \in A(\bar{x})} (\theta_i \kappa(\bar{Y}) \Delta Y_\varepsilon)$$

$$= \alpha \kappa(\bar{Y}) \Delta Y_\varepsilon \sum_{i \in A(\bar{x})} \theta_i$$

$$= \alpha \kappa(\bar{Y}) \Delta Y_\varepsilon$$
where $K = \max_{1 \leq i \leq n} K_i$ and $K_i$ are the local Lipschitz constants of $\nabla g_i$.

Now recall that $|\tilde{n}| = 1$, so

$$\alpha = \left| \sum_{i \in A(\tilde{x})} \theta_i \nabla s g_i (Y) \right| = 1$$

$$\alpha = \frac{1}{\left| \sum_{i \in A(\tilde{x})} \theta_i \nabla s g_i (Y) \right|}$$

$$\alpha \leq \frac{1}{\min_{\theta} \left\{ \left| \sum_{i \in A(\tilde{x})} \theta_i \nabla s g_i (Y) \right| : \theta_i \geq 0, \sum_{i \in A(\tilde{x})} \theta_i = 1 \right\}}$$

$$\alpha \leq \frac{1}{\text{dist}(0, \text{conv} \{ \nabla s g_i (Y) \}_{i \in A(\tilde{x})})}.$$  

Note here that by lemma 2.4 there is no division by 0 in the previous line. By equation (6), we have the following inequality,

$$\frac{2}{\text{dist}(0, \text{conv} \{ \nabla g_i (\tilde{x}) \}_{i \in A(\tilde{x})})} > \frac{1}{\text{dist}(0, \text{conv} \{ \nabla s g_i (Y) \}_{i \in A(\tilde{x})})}.$$  

Applying this inequality, we have

$$\alpha < \frac{2}{\text{dist}(0, \text{conv} \{ \nabla g_i (\tilde{x}) \}_{i \in A(\tilde{x})})}.$$  

Take $K_1 = \frac{2K \kappa(Y)}{\text{dist}(0, \text{conv} \{ \nabla g_i (\tilde{x}) \}_{i \in A(\tilde{x})})}$ and we conclude

$$\text{dist}(\tilde{n}, N_S(\tilde{x})) \leq K_1 \Delta Y.$$  

Now consider $n \in N_S(\tilde{x})$ with $|n| = 1$. Recall that $n \in N_S(\tilde{x})$ implies that

$$n = \beta \sum_{i \in A(\tilde{x})} (\lambda_i \nabla g_i (\tilde{x}))$$

for some $\beta \geq 0, 0 \leq \lambda_i \leq 1, \sum_{i \in A(\tilde{x})} \lambda_i = 1$.

Clearly $n = \beta \sum_{i \in A(\tilde{x})} (\lambda_i \nabla s g_i (Y)) \in \tilde{N}$. Thus we have the following,

$$|\tilde{n} - n| = \left| \beta \sum_{i \in A(\tilde{x})} (\lambda_i \nabla g_i (\tilde{x})) - \beta \sum_{i \in A(\tilde{x})} (\lambda_i \nabla s g_i (Y)) \right|$$

$$\leq \beta \sum_{i \in A(\tilde{x})} |\lambda_i (\nabla g_i (\tilde{x})) - \nabla s g_i (Y)|$$

$$\leq \beta K \kappa(Y) \Delta Y \sum_{i \in A(\tilde{x})} \lambda_i$$

$$= \beta K \kappa(Y) \Delta Y.$$  

Using the same argument as before, noting that $|\tilde{n}| = 1$,

$$\beta \leq \frac{1}{\text{dist}(0, \text{conv} \{ \nabla g_i (\tilde{x}) \}_{i \in A(\tilde{x})})}.$$  

We take $K_2 = \frac{K \kappa(Y)}{\text{dist}(0, \text{conv} \{ \nabla g_i (\tilde{x}) \}_{i \in A(\tilde{x})})}$ and the result follows.

We have now established that the distance between our estimate of the normal cone and the actual normal cone is bounded. We now show with the following theorem that our estimate of the normal cone converges to the actual normal cone.
Theorem 2.8 (convergence to normal cones) Let $S = \{x | g_i(x) \leq 0, i = 1, \ldots, N\}$ where $g_i : \mathbb{R}^n \to \mathbb{R}$. Suppose $\bar{x} \in S$ and LICQ holds at $\bar{x}$. Let $A(\bar{x})$ be the active set of $S$ at $\bar{x}$. Let $Y = \{y_0, \ldots, y_n\}$ form a simplex with $y_0 = \bar{x}$. For $\varepsilon > 0$ let $Y_\varepsilon$ be a simplex defined as in Lemma 1.5 (equation (3)) and define the following set:

$$\tilde{N}_\varepsilon = \text{conv} \{\text{cone} \{\nabla g_i(Y_\varepsilon)\}_i \in A(\bar{x})\}.$$ 

Then

$$\lim_{\varepsilon \to 0} \tilde{N}_\varepsilon = N_S(\bar{x}).$$

(9)

Proof: part I: We first prove the inclusion $\limsup_{\varepsilon \to 0} \tilde{N}_\varepsilon \subseteq N_S(\bar{x})$.

Let $\tilde{n} \in \limsup_{\varepsilon \to 0} \tilde{N}_\varepsilon$. First note that if $\tilde{n} = 0$, then clearly $\tilde{n} \in N_S(\bar{x})$ as $N_S(\bar{x})$ is a cone. Suppose $0 \neq \tilde{n} \in \limsup_{\varepsilon \to 0} \tilde{N}_\varepsilon$. Then there exists $\varepsilon_k \downarrow 0$ and $\tilde{n}_k \in \tilde{N}_{\varepsilon_k}$ such that $\lim_{k \to 0} \tilde{n}_k = \tilde{n}$. Notice that $|\tilde{n}_k| > 0$, so

$$\lim_{k \to 0} \left( \frac{\tilde{n}_k}{|\tilde{n}_k|} \right) = \frac{\tilde{n}}{|\tilde{n}|} \quad \text{with} \quad \left| \frac{\tilde{n}_k}{|\tilde{n}_k|} \right| = 1 \quad \text{and} \quad \left| \frac{\tilde{n}}{|\tilde{n}|} \right| = 1.$$ 

By Theorem 2.7, we know that for each $k$ there exists $n_k \in N_S(\bar{x})$ with

$$\left| \frac{\tilde{n}_k}{|\tilde{n}_k|} - n_k \right| \leq \frac{2K\kappa(Y_{\varepsilon_k}) \Delta Y_{\varepsilon_k}}{\text{dist}(0, \text{conv} \{\nabla g_i(\bar{x})\}_i \in A(\bar{x})).}$$

By Lemma 1.5, we know that $\kappa(Y_{\varepsilon_k}) \Delta Y_{\varepsilon_k} = \varepsilon_k \kappa(Y) \Delta Y$, so

$$0 \leq \left| \frac{\tilde{n}_k}{|\tilde{n}_k|} - n_k \right| \leq \frac{2\varepsilon_k K \kappa(Y) \Delta Y}{\text{dist}(0, \text{conv} \{\nabla g_i(\bar{x})\}_i \in A(\bar{x})).}$$

Combining this with the assumption that $\varepsilon_k \downarrow 0$ implies that $\lim_{k \to 0} n_k = \frac{\tilde{n}}{|\tilde{n}|}$. As $N_S(\bar{x})$ is closed and $n_k \in N_S(\bar{x})$, we have that $\frac{\tilde{n}}{|\tilde{n}|} \in N_S(\bar{x})$. As $N_S(\bar{x})$ is a cone, we have that $\left| \frac{\tilde{n}}{|\tilde{n}|} \right| = \tilde{n} \in N_S(\bar{x})$. Hence $\limsup_{\varepsilon \to 0} \tilde{N}_\varepsilon \subseteq N_S(\bar{x})$.

part II: We now prove the inclusion $\liminf_{\varepsilon \to 0} \tilde{N}_\varepsilon \supseteq N_S(\bar{x})$.

Let $\tilde{n} \in N_S(\bar{x})$. Again, if $\tilde{n} = 0$, then $\tilde{n} \in \tilde{N}_\varepsilon$ for any $\varepsilon > 0$, so $\tilde{n} \in \liminf_{\varepsilon \to 0} \tilde{N}_\varepsilon$. Suppose that $0 \neq \tilde{n} \in N_S(\bar{x})$. Then $\frac{\tilde{n}}{|\tilde{n}|} \in N_S(\bar{x})$. Let

$$\tilde{n} = \alpha \sum_{i \in A(\bar{x})} \theta_i \nabla g_i(\bar{x})$$

with $\alpha \geq 0, 0 \leq \theta_i \leq 1, \sum_{i \in A(\bar{x})} \theta_i = 1$. Let $\varepsilon_\ell \downarrow 0$ be given. Define

$$\tilde{n}_\ell = \alpha \sum_{i \in A(\bar{x})} \theta_i \nabla g_i(Y_{\varepsilon_\ell}).$$

The argument in Theorem 2.7 shows that $\tilde{n}_\ell \in \tilde{N}_{\varepsilon_\ell}$ and

$$\left| \frac{\tilde{n}}{|\tilde{n}|} - \tilde{n}_\ell \right| \leq \frac{K\kappa(Y_{\varepsilon_\ell}) \Delta Y_{\varepsilon_\ell}}{\text{dist}(0, \text{conv} \{\nabla g_i(\bar{x})\}_i \in A(\bar{x})).}$$

8
for all \( \ell \). By Lemma 1.5 we have that
\[
0 \leq \left| \frac{\tilde{n}}{|n|} - \tilde{n}_\ell \right| \leq \frac{K\varepsilon Y(Y) \Delta Y}{\text{dist}(0, \text{conv}\{\nabla g_i(\tilde{x})\}_{i \in A(\tilde{x})})}.
\]
Applying \( \varepsilon \downarrow 0 \), we have \( \lim_{\varepsilon \downarrow 0} \tilde{n}_\ell = \frac{\tilde{n}}{|n|} \) and therefore \( \lim_{\varepsilon \downarrow 0} |\tilde{n}|\tilde{n}_\ell = \tilde{n} \). Clearly \( |\tilde{n}|\tilde{n}_\ell \in \tilde{N}_\ell \) as \( \tilde{N}_\ell \) is a cone.
Therefore \( \lim_{\varepsilon \downarrow 0} \tilde{N}_\varepsilon \supset N_S(\tilde{x}) \).

The inclusions in part I and part II combine to show that \( \lim_{\varepsilon \downarrow 0} \tilde{N}_\varepsilon = N_S(\tilde{x}) \). \( \square \)

### 3 Intersections of Sets

We now shift our focus to intersections of sets and finding approximations for their respective normal cones. We first consider three different cases. The first is the intersection between a finitely constrained set and a set for which the actual normal cone is known. Case two is the intersection between a set generated by an oracle and a set for which the actual normal cone is known. Case three involves the intersection between a finitely constrained set and a set generated by an oracle. Finally, we will give a generalization for the intersection of any number of these three types of sets. To gain an approximation of the normal cones for intersections of sets, we must define the **non-degeneracy condition**. This will be assumed for all sets throughout the rest of this paper.

**Definition 3.1 (non-degeneracy of intersecting sets)** Let \( S_1, S_2, \ldots, S_m \) be closed regular subsets of \( \mathbb{R}^n \). Let \( \tilde{x} \in S_1 \cap S_2 \cap \cdots \cap S_m \). We say that the non-degeneracy condition holds (for \( S_1, S_2, \ldots, S_m \) at \( \tilde{x} \)) if the only combination of vectors \( n_i \in N_{S_i}(\tilde{x}) \) with \( n_1 + \cdots + n_m = 0 \) is \( n_i = 0 \) for all \( i \).

We will now give an example that provides some insight into the non-degeneracy condition. Before we give it, however, we need a proposition that makes use of it.

**Proposition 3.2** Let \( S_1, S_2, \ldots, S_m \) be closed, regular subsets of \( \mathbb{R}^n \). Let \( S = S_1 \cap \cdots \cap S_m \) and let \( \tilde{x} \in S \). Suppose that non-degeneracy condition holds, then \( S \) is regular at \( \tilde{x} \) and
\[
N_S(\tilde{x}) = N_{S_1}(\tilde{x}) + \ldots + N_{S_m}(\tilde{x}) .
\]

**Proof:** See \([RW98, \text{Thm} \ 6.42]\) \( \square \)

This proposition will become imperative to our later results. For now, we use it to prove the following example.

**Example 3.3** Let \( C_1, C_2, \ldots, C_m \) be closed, convex subsets of \( \mathbb{R}^n \). Let \( C = C_1 \cap \cdots \cap C_m \) and let \( \tilde{x} \in C \). Suppose that \( \text{int}(C) \neq \emptyset \). Then the non-degeneracy condition holds for \( C \) at \( \tilde{x} \).

**Proof:** We prove the result via induction. For our base case, consider \( C = C_1 \cap C_2 \). Suppose \( \text{int}(C) \neq \emptyset \). For eventual contradiction, suppose that the non-degeneracy condition does not hold, so we have that \( N_{C_1}(\tilde{x}) \cap (-N_{C_2}(\tilde{x})) \neq \{0\} \). Let \( n_1 \in N_{C_1}(\tilde{x}) \), with \( -n_1 \in N_{C_2}(\tilde{x}) \). Thus we have the following
\[
< n_1, x - \tilde{x} > \leq 0 \quad \text{for all} \quad x \in C_1 \\quad \text{and} \quad \\text{for all} \quad x \in C_2
\]
which implies that
\[
< n_1, x - \tilde{x} > = 0 \quad \text{for all} \quad x \in C_1 \cap C_2.
\]
This gives us that
\( C_1 \cap C_2 \subseteq \{ x | \langle n_1, x - \bar{x} \rangle > 0 \} = H. \)

But \( H \) is a hyperplane, thus \( \text{int}(H) = \emptyset \), which implies that \( \text{int}(C_1 \cap C_2) = \emptyset \) which is a contradiction. Thus the non-degeneracy condition must hold.

Now we assume that if \( \text{int}(C_1 \cap \ldots \cap C_{m-1}) \neq \emptyset \), then the non-degeneracy condition holds for \( C_1 \cap \ldots \cap C_{m-1} \).

Consider \( C = C_1 \cap \ldots \cap C_m \) with \( \text{int}(C) \neq \emptyset \). Let \( T = C_1 \cap \ldots \cap C_{m-1} \). Clearly, as \( \text{int}(C) \neq \emptyset \), we have that \( \text{int}(T) \neq \emptyset \). Thus, by the induction hypothesis, we have that the non-degeneracy condition holds for \( T \).

Thus, as \( C_1, \ldots, C_{m-1} \) are convex, we have by Proposition 3.2

\[
N_T(\bar{x}) = \sum_{i=1}^{m-1} N_C(x).
\]

For eventual contradiction, suppose that the non-degeneracy condition does not hold for \( C \). Then we have that if \( \sum_{i=1}^{m} n_i = 0 \), then \( n_i \neq 0 \) for all \( i \) for \( n_i \in N_C(x) \). Thus we have that \( n_m = -\sum_{i=1}^{m-1} n_i \). But

\[
\sum_{i=1}^{m} n_i \in N_T(\bar{x}) \text{ by Proposition 3.2. Thus we have that } N_T(\bar{x}) \cap (-N_C(x)) \neq \{0\}.
\]

Using the same argument as before, we see that

\[
T \cap C_m \subseteq \{ x | \langle n_m, x - \bar{x} \rangle > 0 \} = H
\]

where \( H \) is a hyperplane, and so \( \text{int}(H) = \emptyset \). But this is a contradiction as \( T \cap C_m = C_1 \cap \ldots \cap C_m = C \) and, by assumption, \( \text{int}(C) \neq \emptyset \).

Thus we have that if \( \text{int}(C) \neq \emptyset \), then the non-degeneracy condition holds for \( C \).

Under the assumption of this non-degeneracy condition, in conjunction with Proposition 3.2, we are able to provide an approximation for the normal cones to sets in case one.

**Corollary 3.4** Let \( S_1 = \{ x | g_i(x) \leq 0, i = 1, \ldots, N \} \) where \( g_i : \mathbb{R}^n \to \mathbb{R} \). Let \( S_2 \) be a closed regular set for which the normal cone is known. Suppose \( \bar{x} \in S = S_1 \cap S_2 \) and LICQ holds at \( \bar{x} \) for \( S_1 \). Let \( A(\bar{x}) \) be the active set of \( S \) at \( \bar{x} \). Let \( Y = \{ y_0, \ldots, y_n \} \) form a simplex with \( y_0 = \bar{x} \). For \( \varepsilon > 0 \) let \( \bar{N}_\varepsilon \) be the set defined in Theorem 2.8. Suppose that the non-degeneracy condition holds, then

\[
\lim_{\varepsilon \downarrow 0} \bar{N}_\varepsilon + N_{S_2}(\bar{x}) = N_{S_1 \cap S_2}(\bar{x}).
\]

**Proof:** Note that as \( g_i \) are smooth, we have that \( S_1 \) is closed and regular. The result then follows from Proposition 3.2 and Theorem 2.8.

In order to find approximations to normal cones of intersections of sets where at least one of the sets is defined by an oracle, we must have a method for finding normal cones to these sets. Fortunately, [HL05] provide us with just the approximation we need.

**Proposition 3.5** Consider a set \( S \subset \mathbb{R}^n \) that is regular at a point \( \bar{x} \in S \). Fix \( \alpha \in (0,1) \). For each \( \varepsilon > 0 \) define the set \( G_\varepsilon \) via

\[
G_\varepsilon = \{ \lambda w | \lambda \geq 0, w = x - P_{\text{conv}(S \cap B_\varepsilon(\bar{x}))}(x), |x - \bar{x}| \leq \varepsilon, |w| \geq \alpha \varepsilon \} \cup \{0\}
\]

where \( P_{\text{conv}(S \cap B_\varepsilon(\bar{x}))}(x) \) denotes the projection of \( x \) onto the set \( \text{conv}\{S \cap B_\varepsilon(\bar{x})\} \). Then

\[
\lim_{\varepsilon \downarrow 0} G_\varepsilon = N_S(\bar{x}).
\]

Furthermore, if \( \text{int}(T_S(\bar{x})) \neq \emptyset \), then

\[
\lim_{\varepsilon \downarrow 0} \text{conv}\{G_\varepsilon\} = N_S(\bar{x}).
\]
Proposition 3.10 (convergence of sums) of limits in case three. To deal with this, we provide the following proposition given by [RW98].

Proof: See [RW98, Thm 4.25]

The following corollary will give us our approximation for normal cones to sets of the type defined by case two.

Corollary 3.6 Let $S_1$ be a closed regular set defined by an oracle. Let $S_2$ be a closed regular set for which the normal cone is known. Let $G_x$ be the set defined in Proposition 3.5. Let $\bar{x} \in S = S_1 \cap S_2$. Suppose that the non-degeneracy condition holds, then

$$\lim_{\varepsilon \downarrow 0} G_x + N_{S_2}(\bar{x}) = N_{S_1 \cap S_2}(\bar{x}).$$

Proof: The result follows directly as an application of Propositions 3.2 and 3.5.

To deal with multiple sets being provided by oracles or a finite number of constraints, we must introduce the notions of horizon cones and total convergence. As usual, we follow the definitions given in [RW98].

Definition 3.7 (horizon cones) For a set $S \subset \mathbb{R}^n$, the horizon cone is the closed cone $S^\infty \subset \mathbb{R}^n$ representing the direction set of $S$:

$$S^\infty = \left\{ \begin{array}{ll} \{x \mid x^v \in S, \lambda^v \downarrow 0, \text{ with } \lambda^v x^v \to x \} & \text{when } S \neq \emptyset, \\ \emptyset & \text{when } S = \emptyset. \end{array} \right.$$  

If $S$ itself happens to be a cone, then $S^\infty = c1S$.

The following definition is equivalent to total set convergence by [RW98, Prop 4.24].

Definition 3.8 (total set convergence) A sequence of sets $S^v \subset \mathbb{R}^n$ is said to converge totally to a closed set $S \subset \mathbb{R}^n$, written $S^v \to^t S$, if $S^v \to S$ and $(S^v)^\infty \to S^\infty$.

Total set convergence will be needed to help us deal with a sum of limits. The next proposition will also prove useful.

Proposition 3.9 If $S^v$ is a cone for all $v$, then $S^v \to S \neq \emptyset$ implies that $S^v \to^t S$.

Proof: See [RW98, Thm 4.25]

As both of our approximations so far have been dependent on the limit of a set, we must consider a sum of limits in case three. To deal with this, we provide the following proposition given by [RW98].

Proposition 3.10 (convergence of sums) If $S^v_1 \to^t S_1, S^v_2 \to^t S_2$, and $S^v_1 \cap (-S^\infty_2) = \{0\}$, then

$$S^v_1 + S^v_2 \to^t S_1 + S_2.$$  

Furthermore, if $S^v_i \to^t S_i, i = 1, \ldots, m, S^\infty_1 \times \ldots \times S^\infty_m = (S_1 \times \ldots \times S_m)^\infty$, and the only combination of vectors $y_i \in S^\infty_i$ with $y_1 + \ldots + y_m = 0$ is $y_i = 0$ for all $i$, then

$$S^v_1 + \ldots + S^v_m \to^t S_1 + \ldots + S_m.$$  

Proof: See [RW98, Exercise 4.29].

The preceding result enables us to provide a result for normal cones to sets defined by case three.

Theorem 3.11 (approximating normal cones of intersections) Let $S_1 = \{x \mid g_i(x) \leq 0, i = 1, \ldots, n\}$ where the $g_i$ are smooth for all $i$. Let $S_2$ be a closed, regular set defined by an oracle. Suppose $\bar{x} \in S = S_1 \cap S_2$ and LICQ holds at $\bar{x}$ for $S_1$. Let $A(\bar{x})$ be the active set of $S$ at $\bar{x}$. Let $Y = \{y_0, \ldots, y_n\}$ form a simplex with $y_0 = \bar{x}$. For $\varepsilon > 0$ let $N_\varepsilon$ be the set defined in Theorem 2.8 and let $G_\varepsilon$ be the set defined in Proposition 3.5. Suppose the non-degeneracy condition holds, then

$$\lim_{\varepsilon \downarrow 0} N_\varepsilon + \lim_{\varepsilon \downarrow 0} G_\varepsilon = \lim_{\varepsilon \downarrow 0}(N_\varepsilon + G_\varepsilon) = N_{S_1 \cap S_2}(\bar{x}).$$

11
Proof: By applying Propositions 3.2 and 3.5 and Theorem 2.8, the following result is clear

\[ N_{S_1 \cap S_2}(\bar{x}) = \lim_{\varepsilon \downarrow 0} \tilde{N}_\varepsilon + \lim_{\varepsilon \downarrow 0} G_\varepsilon. \]

To prove the other identity, we first note that as \( \tilde{N}_\varepsilon \) and \( G_\varepsilon \) are cones for all \( \varepsilon \), we have that \( \tilde{N}_\varepsilon \to^t N_{S_1}(\bar{x}) \) and \( G_\varepsilon \to^t N_{S_2}(\bar{x}) \). Thus it suffices to show that \( N_{S_1}(\bar{x})^\infty \cap (-N_{S_2}(\bar{x})^\infty) = \{0\} \). As \( N_{S_1}(\bar{x}) \) are closed, cones, by 3.7 we have that \( N_{S_1}(\bar{x})^\infty = N_{S_1}(\bar{x}) \). Thus we must have that \( N_{S_1}(\bar{x}) \cap (-N_{S_2}(\bar{x})) = \{0\} \). This is exactly the non-degeneracy condition that we are assuming, therefore Proposition 3.10 holds, and we have that \( \lim_{\varepsilon \downarrow 0} (\tilde{N}_\varepsilon + G_\varepsilon) = \lim_{\varepsilon \downarrow 0} \tilde{N}_\varepsilon + \lim_{\varepsilon \downarrow 0} G_\varepsilon. \)

We are now able to use proposition 3.10 in conjunction with the approximation results for our three cases to derive one unifying theorem for the general case.

Theorem 3.12 Let \( L, O, \) and \( K \) represent three disjoint index sets. For \( k \in L \), let \( S_k = \{x \mid g_i(x) \leq 0, i = 1, \ldots N\} \), where \( g_i : \mathbb{R}^n \to \mathbb{R} \). For \( k \in O, \) let \( S_k \) be closed, regular sets defined by an oracle. For \( k \in K, \) let \( S_k \) be closed, regular sets for which the normal cone is known. Let \( \bar{x} \in \bigcap_{k \in L, k \in O, k \in K} S_k = S, \) and suppose LICQ holds at \( \bar{x} \) for each \( S_k, k \in L \).

For \( \varepsilon > 0 \) and \( k \in L \) let \( N_{S_k}(\bar{x})^\varepsilon \) be the set defined in Theorem 2.8 (using \( Y_\varepsilon \) as generated in Lemma 1.5). For \( \varepsilon > 0 \) and \( k \in O \) let \( G_\varepsilon \) be the set defined in Proposition 3.5. Suppose that the non-degeneracy condition holds, then

\[ N_S(\bar{x}) = \sum_{k \in K} N_{S_k}(\bar{x}) + \sum_{k \in L} \lim_{\varepsilon \downarrow 0} \tilde{N}_\varepsilon + \sum_{k \in O} \lim_{\varepsilon \downarrow 0} G_\varepsilon = \sum_{k \in K} N_{S_k}(\bar{x}) + \lim_{\varepsilon \downarrow 0} \left( \sum_{k \in L} \tilde{N}_\varepsilon + \sum_{k \in O} G_\varepsilon \right) \]

Proof: As the non-degeneracy condition holds, the following result is clear

\[ N_S(\bar{x}) = \sum_{k \in K} N_{S_k}(\bar{x}) + \sum_{k \in L} \lim_{\varepsilon \downarrow 0} \tilde{N}_\varepsilon + \sum_{k \in O} \lim_{\varepsilon \downarrow 0} G_\varepsilon. \]

We need now only show the second equality. From Theorem 3.10, we know that \( \tilde{N}_\varepsilon \to^t N_{S_k}(\bar{x}) \) for all \( k \in L \) and \( G_\varepsilon \to^t N_{S_k}(\bar{x}) \) for all \( k \in O \). Also, \( N_{S_k}(\bar{x}) \) are non-empty, closed, convex sets. Therefore, by [RW98, Pg 127], we have \( \prod_{k \in L, k \in O} N_{S_k}(\bar{x})^\infty = \left( \prod_{k \in L, k \in O} N_{S_k}(\bar{x}) \right)^\infty \). Proposition 3.10 can now be applied so we see that

\[ \sum_{k \in L} \tilde{N}_\varepsilon + \sum_{k \in O} G_\varepsilon \to^t \sum_{k \in L} N_{S_k}(\bar{x}) + \sum_{k \in O} N_{S_k}(\bar{x}). \]

By 3.8 total convergence implies regular convergence, thus we have the following

\[ \lim_{\varepsilon \downarrow 0} \left( \sum_{k \in L} \tilde{N}_\varepsilon + \sum_{k \in O} G_\varepsilon \right) = \sum_{k \in L} N_{S_k}(\bar{x}) + \sum_{k \in O} N_{S_k}(\bar{x}). \]

4 Numerical Tests

In this section we provide numerical results run on a number of sets to compare the efficiency of the simplex gradient in approximating the normal cone to that of the convex hull projection method proposed in [HL05].

We consider three sets for which LICQ holds that exemplify the accuracy of the simplex gradient, and one set for which LICQ fails to see the problems that arise.
4.1 LICQ Satisfied

In this subsection, we consider three families of finitely constrained sets in \( \mathbb{R}^n \) for which LICQ holds:

\[
A_{m,n} = \left\{ x \mid x_n \leq 1, \left( \sum_{i=1}^{n-1} x_i^2 \right)^{m/2} \leq x_n \right\},
\]

(11)

\[
B_{m,n} = \left\{ x \mid x_n \leq \left( \sum_{i=1}^{n-1} (e^{x_i})^2 \right)^{m/2}, M \left( \sum_{i=1}^{n-1} x_i^2 \right)^{m/2} \leq x_n \right\} \quad \text{where} \quad M = \frac{e^2}{(n-1)^{m/2}},
\]

(12)

and

\[
C_{m,n} = \left\{ x \mid \begin{array}{l}
\sum_{i=1}^{k} (k+1-i) \ln(x_i) - 1 \leq 0 \quad \text{when} \quad k \equiv 1, 2(\text{mod} \ 4) \\
\sum_{i=1}^{k} (k+1-i) \ln(x_i) \leq 0 \quad \text{when} \quad k \equiv 0, 3(\text{mod} \ 4)
\end{array},
\right.
\]

(13)

We run tests on all three sets for dimensions \( n = 2, 3, 5, \) and 10. For sets \( A \) and \( B \) we examine \( m = 2, 4, \) and 16. For set \( C \) we examine \( m = 2, 5, \) and 10 when \( m \leq n \). For the set \( A \) we set \( \bar{x} = (1, 0, 0, \ldots, 0, 1) \). For the set \( B \) we set \( \bar{x} = (1, 1, 1, \ldots, 1, (n-1)e^2) \). For the set \( C \) we set \( \bar{x} \) equal to a vector consisting of \( e \) and \( e^{-1} \) where \( e \) is used in the \( i^{th} \) position if \( i \equiv 0, 1(\text{mod} \ 4) \) and \( e^{-1} \) is used otherwise. The correct normal cones to sets \( A, B, \) and \( C \) at their respective \( \bar{x} \) are given by

\[
N_A(\bar{x}) = \text{conv} \{ \{0,0,\ldots,0,1\}, \{(m,0,0,\ldots,0,-1)\} \},
\]

\[
N_B(\bar{x}) = \text{conv} \{ \{me^2, me^2, me^2, \ldots me^2, -1\}, \{-2e^2, -2e^2, \ldots -2e^2, 1\} \},
\]

and

\[
N_C(\bar{x}) = \text{conv} \{ \left\{ \left( \frac{1}{x_1}, 0, 0, \ldots \right), \left( \frac{2}{x_1}, \frac{1}{x_2}, 0, 0, \ldots \right), \ldots, \left( \frac{m}{x_1}, \frac{m-1}{x_2}, \ldots, \frac{1}{x_m}, 0, \ldots \right) \right\} \}.
\]

For each set in each dimension, we test the generating of approximate normal cones using both formula (5) and the numerical formula suggested in [HL05, §6]:

\[
G^e_\lambda = \{ \lambda w \mid \lambda \geq 0, w = x - P_{\text{conv}(SY\varepsilon^2)}(x), |w| \geq \alpha \varepsilon \} \cup \{0\}
\]

(14)

where \( Y^p_\varepsilon \) is a set of \( p \) randomly generated points within \( \varepsilon \) of \( \bar{x} \). For both approaches we test search radii of \( \varepsilon = 1, 0.1, \) and \( 10^{-6} \). For formula (14) we test \( p = n + 1, 10(n + 1), \) and \( 100(n + 1) \).

As in [HL05], the accuracy of the normal cone is calculated by the maximum angle in the difference between the approximated normal cone and the actual normal cone. This is given by the following formula

\[
\max \left\{ \begin{array}{l}
\max\{\text{arcsin}(|w - P_{N_S(\bar{x})}(w)|) : w \in \tilde{N}, |w| = 1\} \\
\max\{\text{arcsin}(|w - P_{S}(w)|) : w \in N_S(\bar{x}), |w| = 1\}
\end{array} \right\}
\]

where \( N_S(\bar{x}) \) is the actual normal cone and \( \tilde{N} \) is the approximate normal cone. To account for the randomness in the techniques, we run each test 10 times. In Tables 3, 4, 5, and 6 (in appendix), we report the mean (standard deviation) of these tests. We summarize the results in Table 1 below. Where a 0 appears in these tables, it means that the result we observed was less than \( 10^{-5} \).

Examining Table 1 and Table 3 we see, as implied by Theorem 2.7, that the accuracy of the generated approximate normal cone improves linearly as \( \varepsilon \) decreases. We also note that even with just \( n + 1 \) points, formula (5) generates a very accurate approximation of the normal cone very quickly. Even in higher dimensions, the simplex gradient technique performs exceedingly well.

Examining Table 1 and Tables 4 to 6, we see that although decreasing \( \varepsilon \) generally improves the quality of the approximation, the impact is not as dramatic. In fact, in some cases the accuracy of the approximations becomes worse as \( \varepsilon \) decreases. As the number of points increase we do see some improvement, but even then once the tests move into higher dimensions or the constraint sets become more complicated, the number of points ceases to matter. We will also note that the computation time needed to run these approximates scales with the number of points, so large numbers of points is impractical for applications.
Table 1: Mean of $N_n$ and $G_{p}^n$ for various tests.

<table>
<thead>
<tr>
<th>Set</th>
<th>$N_1$</th>
<th>$N_{0.1}$</th>
<th>$N_{10^{-6}}$</th>
<th>$G_{1}^{10(n+1)}$</th>
<th>$G_{10^{-6}}^{100(n+1)}$</th>
<th>$G_{0.1}^{100(n+1)}$</th>
<th>$G_{10^{-6}}^{100(n+1)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.6673</td>
<td>0.2785</td>
<td>0</td>
<td>1.2725</td>
<td>1.0796</td>
<td>1.1701</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.5021</td>
<td>0.1302</td>
<td>0</td>
<td>1.2944</td>
<td>1.2676</td>
<td>1.3012</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.7456</td>
<td>0.1637</td>
<td>0</td>
<td>1.0166</td>
<td>1.1184</td>
<td>1.1221</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$G_{10}^{10(n+1)}$</td>
<td>$G_{10^{-6}}^{100(n+1)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1.2114</td>
<td>1.0645</td>
<td>0.9938</td>
<td>1.1496</td>
<td>0.8753</td>
<td>0.8415</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1.4986</td>
<td>1.5079</td>
<td>1.5037</td>
<td>1.3521</td>
<td>1.4064</td>
<td>1.4297</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.3532</td>
<td>1.1441</td>
<td>1.1638</td>
<td>1.3133</td>
<td>1.1168</td>
<td>1.1093</td>
<td></td>
</tr>
</tbody>
</table>

4.2 LICQ Violated

In the tests above we saw great success from the normal cone approximation technique developed in this work. In this subsection, we demonstrate how things can go very wrong when LICQ fails. In particular, we consider a finitely constrained set for which LICQ falters:

$$D = \{ x \mid 1 \leq x_1, |x| \leq 1 \}$$

(15)

at the point $\bar{x} = (1, 0, 0, \ldots, 0)$. We run the tests in dimensions $n = 2, 3, 5,$ and 10. Since the correct normal cone to set $D$ at $\bar{x}$ is $N_D(\bar{x}) = \mathbb{R}^n$.

instead of measuring accuracy via the formula above, we simply examine whether the approximate normal cone is $\mathbb{R}^n$ or is not $\mathbb{R}^n$.

As before we test both approaches for search radii $\varepsilon = 1, 0.1,$ and $10^{-6}$, and test formula 14 for $p = n + 1, 10(n + 1),$ and $100(n + 1)$. To account for the randomness in the techniques, we run each test 1000 times. In Table 2 we report the percentage of time the approximate normal cone was correct.

Table 2: Percentage of tests where the approximate normal cone was correct.

<table>
<thead>
<tr>
<th>n</th>
<th>$\varepsilon$</th>
<th>$N_{\varepsilon}$</th>
<th>$G_{\varepsilon}^{(n+1)}$</th>
<th>$G_{\varepsilon}^{10(n+1)}$</th>
<th>$G_{\varepsilon}^{100(n+1)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1 0%</td>
<td>22.4%</td>
<td>100%</td>
<td>99.9%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.1 0%</td>
<td>21.6%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$10^{-6}$ 0%</td>
<td>22.2%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
<td>10.9%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.1 0%</td>
<td>11.8%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$10^{-6}$ 0%</td>
<td>12.7%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>2.7%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.1 0%</td>
<td>3.0%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$10^{-6}$ 0%</td>
<td>2.0%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.1 0%</td>
<td>0.1%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$10^{-6}$ 0%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Examining Table 2 it is clear that when LICQ fails, the approximation technique developed in this paper is unsuccessful. This corresponds with the theory as the LICQ assumption is not met. In this example, because the LICQ assumption is not met, the simplex gradient approach is not provided with enough information to approximate the normal cone. In particular, this approach will approximate the normal cone as the conic hull of two vectors, this can never generate $\mathbb{R}^n$.
Conversely, for this set, the technique in formula (14) is quite reasonable as soon as \( p \) is sufficiently large. In this case, formula (14) will generate the correct normal cone provided the random set \( Y_p \) provides reasonable coverage of \( \mathbb{R}^n \) in the sense that \( 0 \in \text{int} (\text{conv} (Y_p)) \). As \( p \) increases this becomes increasingly likely. (A probabilistic analysis of when this occurs is likely achievable, but beyond the scope of this paper.)

5 Conclusion

In this paper we have examined the challenge of generating a normal cone to a set without analytic knowledge of the set. Research in this area began in [HL05], where a method of approximating a normal cone to an oracle based set based on projections was developed. In this paper, we began by considering sets given by the form

\[
S = \{ x \mid g_i(x) \leq 0, \ i = 1, \ldots, N \},
\]

where each \( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \) is unknown and provided by an oracle (or black-box). Introducing the idea of a simplex gradient, we proved that under a linear independence constraint qualification it was possible to build an approximate normal cone by taking the convex conic hull of the active simplex gradients (see Theorems 2.7 and 2.8). We then considered the situation when

\[
S = S_1 \cap S_2 \cap \ldots \cap S_m,
\]

for sets \( S_k \) of various forms. In Theorem 3.11 it is shown that, under a non-degeneracy condition, approximating normal cones to intersections of sets is possible by taking sums of approximations. This result holds both for the simplex gradient based approximation (developed herein) and for the convex hull projection method proposed in [HL05]. Indeed, it is even possible to generate approximate normals for different sets using different techniques.

In Section 4 we provided several examples demonstrating the increased accuracy that is obtainable by applying the simplex gradient based approximation over the convex hull projection method. We also provide one example of how the simplex gradient based approximation can fail when LICQ is not satisfied.

Although in Section 3 we focused on the question of finding normal cones to intersections of sets, one could easily consider extensions of these ideas in the direction of sums of sets via results like the following (see [RW98, Exercise 6.42]).

**Proposition 5.1 (normals under set addition)** Let \( C = C_1 + \ldots + C_m \) for closed convex sets \( C_i \in \mathbb{R}^n \). Then for any choice of \( \bar{x}_i \in C_i \) with \( \bar{x}_1 + \ldots + \bar{x}_m = \bar{x} \), one has that

\[
N_C(\bar{x}) = N_{C_1}(\bar{x}_1) \cap \ldots \cap N_{C_m}(\bar{x}_m).
\]

It is also worth noting that we present this research in the framework of normal cones, as opposed to tangent cones. The tangent cone, \( T_S(x) \), can be derived with relative ease from our normal cone approximation. We refer to the following proposition in [RW98].

**Proposition 5.2 (tangent-normal polarity)** Let \( S = \{ x \mid g_i(x) \leq 0 \} \subset \mathbb{R}^n \) be locally closed at \( \bar{x} \in S \). Then

\[
N_S(\bar{x}) = T_S(\bar{x})^*,
\]

where \( K^* \) denotes the polar cone to the cone \( K : K^* = \{ y \in \mathbb{R}^n \mid \langle x, y \rangle \leq 0 \text{ for all } x \in K \} \).

References


Table 3: Mean and standard deviation of $N_c$ test results after 10 iterations.

<table>
<thead>
<tr>
<th>Set</th>
<th>$N_1$</th>
<th>$N_{0.1}$</th>
<th>$N_{10^{-6}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{2,2}$</td>
<td>0.0406 (0.1063)</td>
<td>0.0053 (0.0169)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{4,2}$</td>
<td>0.0390 (0.1234)</td>
<td>0.0493 (0.0653)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{16,2}$</td>
<td>0.4160 (0.2459)</td>
<td>0.1481 (0.1481)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{2,3}$</td>
<td>0.4029 (0.3390)</td>
<td>0.0518 (0.0310)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{4,3}$</td>
<td>0.2914 (0.1968)</td>
<td>0.0929 (0.1482)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{16,3}$</td>
<td>0.8181 (0.3957)</td>
<td>0.4083 (0.4701)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{2,5}$</td>
<td>0.6029 (0.2444)</td>
<td>0.0712 (0.0366)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{4,5}$</td>
<td>0.9123 (0.4176)</td>
<td>0.3186 (0.4228)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{16,5}$</td>
<td>1.0409 (0.4530)</td>
<td>0.5112 (0.3869)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{2,10}$</td>
<td>0.8664 (0.1844)</td>
<td>0.2840 (0.2954)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{4,10}$</td>
<td>1.2617 (0.3057)</td>
<td>0.5595 (0.5034)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$A_{16,10}$</td>
<td>1.2957 (0.2586)</td>
<td>0.8240 (0.2622)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{2,2}$</td>
<td>0.0170 (0.0083)</td>
<td>0.0039 (0.0014)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{4,2}$</td>
<td>0.0234 (0.0103)</td>
<td>0.0033 (0.0008)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{16,2}$</td>
<td>0.0474 (0.0281)</td>
<td>0.0148 (0.0135)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{2,3}$</td>
<td>0.4508 (0.4813)</td>
<td>0.0173 (0.0107)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{4,3}$</td>
<td>0.2136 (0.1549)</td>
<td>0.0384 (0.0552)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{16,3}$</td>
<td>0.4574 (0.4679)</td>
<td>0.1375 (0.1318)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{2,5}$</td>
<td>0.6739 (0.3943)</td>
<td>0.0796 (0.0371)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{4,5}$</td>
<td>0.515 (0.2433)</td>
<td>0.0613 (0.0382)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{16,5}$</td>
<td>0.9339 (0.3734)</td>
<td>0.3678 (0.4743)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{2,10}$</td>
<td>0.7748 (0.3184)</td>
<td>0.3203 (0.4443)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{4,10}$</td>
<td>0.8220 (0.3534)</td>
<td>0.1395 (0.1084)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$B_{16,10}$</td>
<td>1.0694 (0.3599)</td>
<td>0.3784 (0.2816)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{2,2}$</td>
<td>0.467 (0.5511)</td>
<td>0.0707 (0.0860)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{2,3}$</td>
<td>0.4666 (0.4433)</td>
<td>0.0867 (0.0616)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{2,5}$</td>
<td>0.7651 (0.4610)</td>
<td>0.1150 (0.0722)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{2,10}$</td>
<td>0.6421 (0.2954)</td>
<td>0.2645 (0.3534)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{5,5}$</td>
<td>0.9123 (0.2632)</td>
<td>0.2278 (0.4748)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{5,10}$</td>
<td>1.2283 (0.2810)</td>
<td>0.1865 (0.1426)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$C_{10,10}$</td>
<td>1.0348 (0.1771)</td>
<td>0.195 (0.0900)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Table 4: Mean and standard deviation of $G^{n+1}$ test results after 10 iterations.

<table>
<thead>
<tr>
<th>Set</th>
<th>$G_1^{n+1}$</th>
<th>$G_0^{n+1}$</th>
<th>$G_{10}^{n+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{2,2}$</td>
<td>0.7888 (0.4466)</td>
<td>0.6487 (0.5682)</td>
<td>0.6624 (0.5232)</td>
</tr>
<tr>
<td>$A_{4,2}$</td>
<td>0.9443 (0.6822)</td>
<td>0.7945 (0.5611)</td>
<td>0.8870 (0.6887)</td>
</tr>
<tr>
<td>$A_{16,2}$</td>
<td>0.7075 (0.5580)</td>
<td>0.5433 (0.4503)</td>
<td>0.5252 (0.5110)</td>
</tr>
<tr>
<td>$A_{2,3}$</td>
<td>1.2428 (0.2866)</td>
<td>1.2913 (0.2268)</td>
<td>1.1638 (0.1699)</td>
</tr>
<tr>
<td>$A_{4,3}$</td>
<td>1.1833 (0.3920)</td>
<td>1.2279 (0.3097)</td>
<td>1.1808 (0.2811)</td>
</tr>
<tr>
<td>$A_{16,3}$</td>
<td>1.3014 (0.3283)</td>
<td>1.2352 (0.2640)</td>
<td>1.2366 (0.3274)</td>
</tr>
<tr>
<td>$A_{2,4}$</td>
<td>1.4654 (0.1488)</td>
<td>1.3622 (0.1669)</td>
<td>1.3389 (0.3132)</td>
</tr>
<tr>
<td>$A_{4,4}$</td>
<td>1.4694 (0.1438)</td>
<td>1.4119 (0.1908)</td>
<td>1.2575 (0.2794)</td>
</tr>
<tr>
<td>$A_{16,4}$</td>
<td>1.4762 (0.2256)</td>
<td>1.2938 (0.1475)</td>
<td>1.4163 (0.1230)</td>
</tr>
<tr>
<td>$A_{2,5}$</td>
<td>1.5654 (0.0152)</td>
<td>1.4538 (0.0942)</td>
<td>1.4724 (0.0494)</td>
</tr>
<tr>
<td>$A_{4,5}$</td>
<td>1.5708 (0)</td>
<td>1.4792 (0.0863)</td>
<td>1.4835 (0.0865)</td>
</tr>
<tr>
<td>$A_{16,5}$</td>
<td>1.5545 (0.0313)</td>
<td>1.4487 (0.1248)</td>
<td>1.4263 (0.1119)</td>
</tr>
<tr>
<td>$B_{2,2}$</td>
<td>1.1616 (0.1716)</td>
<td>1.1618 (0.2548)</td>
<td>1.0266 (0.3444)</td>
</tr>
<tr>
<td>$B_{2,4}$</td>
<td>0.6091 (0.5972)</td>
<td>0.5920 (0.5616)</td>
<td>1.0016 (0.4744)</td>
</tr>
<tr>
<td>$B_{16,2}$</td>
<td>0.8332 (0.6351)</td>
<td>0.8750 (0.5287)</td>
<td>0.7691 (0.4177)</td>
</tr>
<tr>
<td>$B_{2,3}$</td>
<td>1.4136 (0.1983)</td>
<td>1.3586 (0.1798)</td>
<td>1.4157 (0.0728)</td>
</tr>
<tr>
<td>$B_{4,3}$</td>
<td>1.3844 (0.1654)</td>
<td>1.2115 (0.2953)</td>
<td>1.3901 (0.2323)</td>
</tr>
<tr>
<td>$B_{16,3}$</td>
<td>1.3007 (0.2114)</td>
<td>1.1467 (0.2724)</td>
<td>1.1426 (0.3284)</td>
</tr>
<tr>
<td>$B_{2,5}$</td>
<td>1.4850 (0.1075)</td>
<td>1.4592 (0.0802)</td>
<td>1.4642 (0.0774)</td>
</tr>
<tr>
<td>$B_{4,5}$</td>
<td>1.4155 (0.1027)</td>
<td>1.4816 (0.0721)</td>
<td>1.4362 (0.0809)</td>
</tr>
<tr>
<td>$B_{16,5}$</td>
<td>1.3971 (0.1469)</td>
<td>1.4027 (0.1570)</td>
<td>1.4425 (0.0722)</td>
</tr>
<tr>
<td>$B_{2,10}$</td>
<td>1.5339 (0.0352)</td>
<td>1.5381 (0.0202)</td>
<td>1.5504 (0.0133)</td>
</tr>
<tr>
<td>$B_{4,10}$</td>
<td>1.4864 (0.0702)</td>
<td>1.4734 (0.0589)</td>
<td>1.5206 (0.0286)</td>
</tr>
<tr>
<td>$B_{16,10}$</td>
<td>1.5125 (0.0391)</td>
<td>1.5131 (0.0445)</td>
<td>1.5150 (0.0413)</td>
</tr>
<tr>
<td>$C_{2,2}$</td>
<td>0.2451 (0.4299)</td>
<td>0.5408 (0.5993)</td>
<td>0.2435 (0.4900)</td>
</tr>
<tr>
<td>$C_{2,3}$</td>
<td>0.7615 (0.6347)</td>
<td>0.9340 (0.3279)</td>
<td>1.1199 (0.5082)</td>
</tr>
<tr>
<td>$C_{2,5}$</td>
<td>1.3320 (0.2260)</td>
<td>1.1883 (0.3534)</td>
<td>1.2886 (0.2496)</td>
</tr>
<tr>
<td>$C_{2,10}$</td>
<td>1.3784 (0.1052)</td>
<td>1.5099 (0.0795)</td>
<td>1.4398 (0.1435)</td>
</tr>
<tr>
<td>$C_{5,5}$</td>
<td>0.9472 (0.2391)</td>
<td>1.0147 (0.2636)</td>
<td>1.1052 (0.2133)</td>
</tr>
<tr>
<td>$C_{5,10}$</td>
<td>1.1939 (0.2227)</td>
<td>1.3497 (0.1662)</td>
<td>1.3886 (0.1668)</td>
</tr>
<tr>
<td>$C_{10,10}$</td>
<td>1.2580 (0.1532)</td>
<td>1.2911 (0.1311)</td>
<td>1.2692 (0.1556)</td>
</tr>
</tbody>
</table>
Table 5: Mean and standard deviation of $G_x^{10(n+1)}$ test results after 10 iterations.

<table>
<thead>
<tr>
<th>Set</th>
<th>$G_1^{10(n+1)}$</th>
<th>$G_{0.1}^{10(n+1)}$</th>
<th>$G_{10^{-6}}^{10(n+1)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{2,2}$</td>
<td>0.4692 (0.3588)</td>
<td>0.7932 (0.2378)</td>
<td>0.4111 (0.1835)</td>
</tr>
<tr>
<td>$A_{4,2}$</td>
<td>0.5557 (0.2963)</td>
<td>0.5592 (0.3299)</td>
<td>0.5939 (0.3458)</td>
</tr>
<tr>
<td>$A_{16,2}$</td>
<td>0.8892 (0.3310)</td>
<td>0.5387 (0.1714)</td>
<td>0.3339 (0.2363)</td>
</tr>
<tr>
<td>$A_{2,3}$</td>
<td>1.2542 (0.3182)</td>
<td>1.0329 (0.1626)</td>
<td>1.0574 (0.2314)</td>
</tr>
<tr>
<td>$A_{4,3}$</td>
<td>1.1037 (0.2889)</td>
<td>1.0787 (0.3488)</td>
<td>0.9959 (0.2356)</td>
</tr>
<tr>
<td>$A_{16,3}$</td>
<td>1.2594 (0.1960)</td>
<td>1.0101 (0.1970)</td>
<td>0.8396 (0.1662)</td>
</tr>
<tr>
<td>$A_{2,5}$</td>
<td>1.4580 (0.0931)</td>
<td>1.3033 (0.1315)</td>
<td>1.2110 (0.1550)</td>
</tr>
<tr>
<td>$A_{4,5}$</td>
<td>1.4461 (0.0926)</td>
<td>1.2253 (0.1338)</td>
<td>1.2560 (0.0828)</td>
</tr>
<tr>
<td>$A_{16,5}$</td>
<td>1.4252 (0.1208)</td>
<td>1.1838 (0.0920)</td>
<td>1.1778 (0.1174)</td>
</tr>
<tr>
<td>$A_{2,10}$</td>
<td>1.5708 (0)</td>
<td>1.3674 (0.0661)</td>
<td>1.3683 (0.0579)</td>
</tr>
<tr>
<td>$A_{4,10}$</td>
<td>1.5685 (0.0074)</td>
<td>1.3593 (0.0687)</td>
<td>1.3423 (0.0542)</td>
</tr>
<tr>
<td>$A_{16,10}$</td>
<td>1.5367 (0.0501)</td>
<td>1.3241 (0.0649)</td>
<td>1.3381 (0.0567)</td>
</tr>
<tr>
<td>$B_{2,2}$</td>
<td>1.5214 (0.0260)</td>
<td>1.5109 (0.0926)</td>
<td>1.5066 (0.0591)</td>
</tr>
<tr>
<td>$B_{4,2}$</td>
<td>1.3208 (0.4390)</td>
<td>1.3017 (0.4581)</td>
<td>1.1641 (0.6083)</td>
</tr>
<tr>
<td>$B_{16,2}$</td>
<td>1.2992 (0.3783)</td>
<td>1.3823 (0.1928)</td>
<td>1.4885 (0.0748)</td>
</tr>
<tr>
<td>$B_{2,3}$</td>
<td>1.5450 (0.0274)</td>
<td>1.5460 (0.0194)</td>
<td>1.5384 (0.0252)</td>
</tr>
<tr>
<td>$B_{4,3}$</td>
<td>1.5830 (0.0385)</td>
<td>1.5072 (0.0597)</td>
<td>1.5011 (0.0620)</td>
</tr>
<tr>
<td>$B_{16,3}$</td>
<td>1.4988 (0.0781)</td>
<td>1.5113 (0.0898)</td>
<td>1.4924 (0.1177)</td>
</tr>
<tr>
<td>$B_{2,5}$</td>
<td>1.5489 (0.0260)</td>
<td>1.5603 (0.0099)</td>
<td>1.5529 (0.0158)</td>
</tr>
<tr>
<td>$B_{4,5}$</td>
<td>1.5332 (0.0421)</td>
<td>1.5469 (0.0206)</td>
<td>1.5547 (0.0063)</td>
</tr>
<tr>
<td>$B_{16,5}$</td>
<td>1.5043 (0.0690)</td>
<td>1.5337 (0.0212)</td>
<td>1.5526 (0.0142)</td>
</tr>
<tr>
<td>$B_{2,10}$</td>
<td>1.5524 (0.0195)</td>
<td>1.5664 (0.0038)</td>
<td>1.5674 (0.0029)</td>
</tr>
<tr>
<td>$B_{4,10}$</td>
<td>1.5350 (0.0354)</td>
<td>1.5615 (0.0061)</td>
<td>1.5625 (0.0081)</td>
</tr>
<tr>
<td>$B_{16,10}$</td>
<td>1.5431 (0.0290)</td>
<td>1.5666 (0.0055)</td>
<td>1.5627 (0.0067)</td>
</tr>
<tr>
<td>$C_{2,2}$</td>
<td>0.8463 (0.5866)</td>
<td>0.4058 (0.2891)</td>
<td>0.4499 (0.3382)</td>
</tr>
<tr>
<td>$C_{2,3}$</td>
<td>1.4679 (0.1466)</td>
<td>0.9962 (0.3200)</td>
<td>1.0107 (0.3047)</td>
</tr>
<tr>
<td>$C_{2,5}$</td>
<td>1.3764 (0.2209)</td>
<td>1.3497 (0.1439)</td>
<td>1.3386 (0.1407)</td>
</tr>
<tr>
<td>$C_{2,10}$</td>
<td>1.5559 (0.0319)</td>
<td>1.5622 (0.0233)</td>
<td>1.5627 (0.0255)</td>
</tr>
<tr>
<td>$C_{5,5}$</td>
<td>1.3254 (0.1649)</td>
<td>0.9962 (0.1173)</td>
<td>1.0794 (0.1754)</td>
</tr>
<tr>
<td>$C_{5,10}$</td>
<td>1.4820 (0.0976)</td>
<td>1.4480 (0.0983)</td>
<td>1.4586 (0.0726)</td>
</tr>
<tr>
<td>$C_{10,10}$</td>
<td>1.4188 (0.0806)</td>
<td>1.2504 (0.0407)</td>
<td>1.2469 (0.0659)</td>
</tr>
</tbody>
</table>
Table 6: Mean and standard deviation of $G_{100(\text{n+1})}$ test results after ten iterations.

<table>
<thead>
<tr>
<th>Set</th>
<th>$G_{100(\text{n+1})}$</th>
<th>$G_{0.1}$</th>
<th>$G_{10^{-6}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_{2,2}</td>
<td>0.4390 (0.1891)</td>
<td>0.2503 (0.1260)</td>
<td>0.2707 (0.2078)</td>
</tr>
<tr>
<td>A_{4,2}</td>
<td>0.5674 (0.2963)</td>
<td>0.3407 (0.1732)</td>
<td>0.2188 (0.1113)</td>
</tr>
<tr>
<td>A_{16,2}</td>
<td>0.3643 (0.2465)</td>
<td>0.2483 (0.1091)</td>
<td>0.2132 (0.3125)</td>
</tr>
<tr>
<td>A_{2,3}</td>
<td>1.2515 (0.2244)</td>
<td>0.7362 (0.1596)</td>
<td>0.6356 (0.1471)</td>
</tr>
<tr>
<td>A_{4,3}</td>
<td>1.1381 (0.1948)</td>
<td>0.7051 (0.1477)</td>
<td>0.6268 (0.1527)</td>
</tr>
<tr>
<td>A_{2,5}</td>
<td>1.0682 (0.2703)</td>
<td>0.6503 (0.1493)</td>
<td>0.7337 (0.1128)</td>
</tr>
<tr>
<td>A_{4,5}</td>
<td>1.4627 (0.1134)</td>
<td>1.1494 (0.1337)</td>
<td>1.1279 (0.0953)</td>
</tr>
<tr>
<td>A_{16,5}</td>
<td>1.4169 (0.1331)</td>
<td>1.1588 (0.0766)</td>
<td>1.1158 (0.0753)</td>
</tr>
<tr>
<td>A_{2,10}</td>
<td>1.4419 (0.0913)</td>
<td>1.1393 (0.0630)</td>
<td>1.0703 (0.0965)</td>
</tr>
<tr>
<td>A_{4,10}</td>
<td>1.5400 (0.0366)</td>
<td>1.3922 (0.0302)</td>
<td>1.3956 (0.0380)</td>
</tr>
<tr>
<td>A_{16,10}</td>
<td>1.5583 (0.0196)</td>
<td>1.3704 (0.0362)</td>
<td>1.3635 (0.0364)</td>
</tr>
<tr>
<td>B_{2,2}</td>
<td>1.5066 (0.0068)</td>
<td>1.5638 (0.0077)</td>
<td>1.5672 (0.0032)</td>
</tr>
<tr>
<td>B_{4,2}</td>
<td>0.7087 (0.7252)</td>
<td>0.7346 (0.7071)</td>
<td>0.6821 (0.7574)</td>
</tr>
<tr>
<td>B_{16,2}</td>
<td>0.5248 (0.6282)</td>
<td>0.6860 (0.6560)</td>
<td>0.8651 (0.6597)</td>
</tr>
<tr>
<td>B_{2,3}</td>
<td>1.4775 (0.1386)</td>
<td>1.5625 (0.0150)</td>
<td>1.5689 (0.0024)</td>
</tr>
<tr>
<td>B_{4,3}</td>
<td>1.3496 (0.2110)</td>
<td>1.4493 (0.1437)</td>
<td>1.5353 (0.0549)</td>
</tr>
<tr>
<td>B_{16,3}</td>
<td>1.3932 (0.1460)</td>
<td>1.5114 (0.0686)</td>
<td>1.5195 (0.0799)</td>
</tr>
<tr>
<td>B_{2,5}</td>
<td>1.5171 (0.1127)</td>
<td>1.5654 (0.0053)</td>
<td>1.5690 (0.0016)</td>
</tr>
<tr>
<td>B_{4,5}</td>
<td>1.5230 (0.0530)</td>
<td>1.5563 (0.0186)</td>
<td>1.5637 (0.0080)</td>
</tr>
<tr>
<td>B_{16,5}</td>
<td>1.4978 (0.0689)</td>
<td>1.5473 (0.0266)</td>
<td>1.5577 (0.0111)</td>
</tr>
<tr>
<td>B_{2,10}</td>
<td>1.5610 (0.0086)</td>
<td>1.5679 (0.0024)</td>
<td>1.5705 (0.0003)</td>
</tr>
<tr>
<td>B_{4,10}</td>
<td>1.5548 (0.0124)</td>
<td>1.5674 (0.0043)</td>
<td>1.5693 (0.0012)</td>
</tr>
<tr>
<td>B_{16,10}</td>
<td>1.5566 (0.0105)</td>
<td>1.5648 (0.0065)</td>
<td>1.5671 (0.0032)</td>
</tr>
<tr>
<td>C_{2,2}</td>
<td>0.8037 (0.6541)</td>
<td>0.2629 (0.2361)</td>
<td>0.1406 (0.1706)</td>
</tr>
<tr>
<td>C_{2,3}</td>
<td>1.0008 (0.1846)</td>
<td>0.7573 (0.1545)</td>
<td>0.7955 (0.1141)</td>
</tr>
<tr>
<td>C_{2,5}</td>
<td>1.3904 (0.1330)</td>
<td>1.3232 (0.1083)</td>
<td>1.3662 (0.1271)</td>
</tr>
<tr>
<td>C_{2,10}</td>
<td>1.5708 (0)</td>
<td>1.5708 (0)</td>
<td>1.5708 (0)</td>
</tr>
<tr>
<td>C_{5,5}</td>
<td>1.2958 (0.2150)</td>
<td>1.0565 (0.0974)</td>
<td>1.0354 (0.0877)</td>
</tr>
<tr>
<td>C_{5,10}</td>
<td>1.5708 (0)</td>
<td>1.5419 (0.0377)</td>
<td>1.5531 (0.0225)</td>
</tr>
<tr>
<td>C_{10,10}</td>
<td>1.5617 (0.0288)</td>
<td>1.3051 (0.0333)</td>
<td>1.3037 (0.0355)</td>
</tr>
</tbody>
</table>