International Journal of Computational Geometry & Applications

Nonsmooth Analysis, Convex Analysis, and their Applications to Motion Planning

Howie Choset
Carnegie Mellon University
Scaife Hall
Pittsburgh, PA 15213

ABSTRACT

Nonsmooth analysis of a broad class of functions taking the form $F(x) = \min_i f_i(x)$, where each f_i is a convex function. One element of this class of functions is the distance function, which measures the distance between a point and the nearest point on the nearest obstacle. Many motion planning algorithms are based on the distance function, and thus rigorous analysis of the distance function can provide a better understanding of how to implement traditional motion planning algorithms. Finally, this paper enumerates some useful results in convex analysis.

Keywords: Nonsmooth analysis, convex analysis, motion planning, roadmaps, Voronoi diagrams.

1. Introduction

Robotic motion planning determines a path between two points q_{start} and q_{goal}, while avoiding obstacles $\{C_i : i = 1, \ldots, n\}$. This one-dimensional path may exist in the robot's work space or in the robot's *configuration space,* the set of all robot locations and postures in a particular work space. Let $\mathcal{W}$ denote the robot's work space or configuration space. In actuality, the path exists in the free space, $\mathcal{FS} = \mathcal{W} \backslash \bigcup_{i=1}^{n} C_i$.

There is a vast literature on *complete* motion planning algorithms for when full knowledge of the world is available to the robot [2]. Two classes of such algorithms are potential field approaches and roadmap methods [2]. In the potential field approach [2,10], the robot is modeled as a particle acting under the influence of a potential function $U \colon \mathcal{FS} \to \mathbb{R}$ that encodes information about the environment such as the goal and obstacle location. Motion planning is effected by gradient descent of U. In order to determine the gradient, these methods assume that U is smooth, but nonsmooth potential functions are often considered.

A roadmap [1,2,3] is a collection of one-dimensional curves that capture the important topological and geometric properties of a robot's environment, i.e.,

Definition 1 (Roadmap) *A* ***roadmap*** *R is the union of one-dimensional curves such that for all q_{start} and q_{goal} in $\mathcal{FS}$, there exists a path between q_{start} and q_{goal} if and only if*

1. *there exists a path from $q_{start} \in \mathcal{FS}$ to some $q'_{start} \in R$* ***(accessibility)***,

2. there exists a path from $q_{goal} \in \mathcal{FS}$ *to some* $q'_{goal} \in R$ ***(departability)****, and*

3. there exists a path in R *between* q'_{start} *and* q'_{goal} ***(connectivity)****.*

Using a roadmap, a planner can construct a path between any two points in a connected component of the robot's free space by first finding a collision free path onto the roadmap (accessibility), traversing the roadmap to the vicinity of the goal (connectivity), and then constructing a collision free path from a point on the roadmap to the goal (departability).

Like potential functions, many roadmaps (Section 2) exploit the differential properties of functions that are not necessarily guaranteed to be smooth. One such function is the distance function, defined in Section 3, that measures the distance between a point and its *closest* obstacle.

In Section 3, it is shown that the distance function is indeed nonsmooth and thus does not posses a conventional gradient. Based on previous results in nonsmooth analysis, summarized in Section 4, Section 5 derives the calculus to differentiate a class of nonsmooth functions of which the distance function is a member. This calculus is then applied in Section 6 to the roadmaps first introduced in Section 2. It is worth noting that the derivations in Section 5 require some new results in convex analysis, described in the Appendix.

2. Motion Planning Motivation

Motion planning using potential fields and roadmap methods provided the initial motivation for this work. This section briefly describes two roadmaps, the opportunistic path planner (OPP) and the generalized Voronoi graph (GVG). The OPP is defined in terms of a nonsmooth function. The GVG is defined in terms of a smooth function, but its accessibility criterion uses the same nonsmooth function which defines the OPP. Finally, using nonsmooth analysis, we show how the procedure to generate the GVG can be applied to the OPP; this enables curve tracing of nonsmooth extrema.

2.1. Opportunistic Path Planner

The *Opportunistic Path Planner* (OPP) [4] is a roadmap method which traces local maxima of a nonsmooth potential function, restricted to a slice, as the slice sweeps through the workspace or a parameterization of the configuration space. A slice in this context is co-dimension one hyperplane and the slice direction is the vector orthogonal to the hyperplane. Note, the slice direction is fixed. Canny and Lin suggest that a distance function evaluated on the slice be used as the potential function. Local maxima on the slice of the distance function are points on the OPP roadmap. The traces of the local maxima as the slice is swept through the work space or configuration space, are termed *freeways*.

First, a fixed slice direction is chosen. The algorithm initially traces a path from the start to the freeway by performing gradient ascent on the distance function in the slice that contains the start. Likewise, a path is traced from the goal to

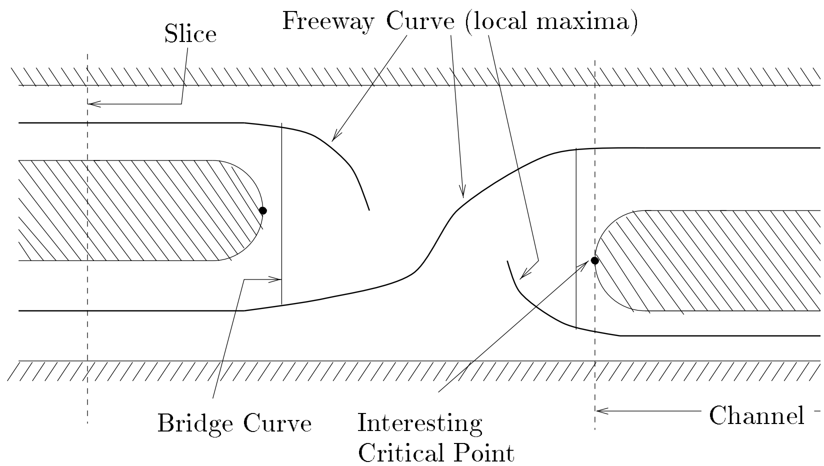

Figure 1: Schematic of the OPP planning scheme.

the freeway via slice-constrained gradient ascent. These two actions correspond to accessibility and departability, as described above.

From the point where the robot accessed the OPP, the algorithm sweeps a slice through $\mathcal{W}$ tracing local maxima of D constrained to the slice, i.e., the algorithm constructs a freeway. If the start and goal freeways are connected, then the algorithm terminates. In general, the set of freeways will not be connected, and paths between neighboring freeways must be found. The OPP freeways are connected via *bridge curves.* The bridge curves are constructed in the vicinity of *interesting critical points.* Interesting critical points occur when *channels* (Figure 1) join or split. This procedure is repeated until the start and goal freeway curves are connected, or all interesting critical points have been explored in which case there does not exist a path between the start and the goal. The union of bridge and freeway curves, sometimes termed a *skeleton*, forms the one-dimensional roadmap.

As will be shown in Section 3, the distance function is nonsmooth, so conventional calculus cannot determine local maxima of the distance. Canny and Lin do not address the issues of nonsmoothness in defining their OPP freeways using the distance function. Without addressing this point, constructing the OPP freeways is not possible in higher dimensioned workspaces. Nevertheless, founded in the basics outlined in Section 4, the results of Section 5 furnish an exact method for determining these local maxima.

2.2. Accessibility of the Generalized Voronoi Graph

Another roadmap method is the *hierarchical generalized Voronoi graph* (HGVG) which is based on the *generalized Voronoi graph* (GVG). The GVG edges are the locus of points equidistant to m obstacles in m dimensions such that the m obstacles are the closest obstacles [6]. In the planar case ($m = 2$), the GVG reduces to the *generalized Voronoi diagram* [5].

Recall that accessibility is the property that a path can be constructed from any point in the free space to a point on the GVG. This is achieved by a cascading sequence of gradient ascent operations of the distance function. First, the robot increases its distance from the nearest obstacle until it is equidistant to two obstacles. Then, maintaining double equidistance, the robot increases distance until it is equidistant to three obstacles. This procedure is repeated until the robot is equidistant to m obstacles. Based on the results of Section 4, Section 6.2 guarantees the accessibility path reaches the GVG from any point in the freespace.

2.3. Incremental Construction of Roadmaps

A sensor based implementation of a roadmap method requires an incremental construction procedure that enables the robot to simultaneously move through the environment and build the roadmap. A key feature of the GVG is that it can be incrementally constructed using line of sight distance information [7], but the OPP method does not have such a method. This paper shows how the OPP method can use the GVG construction procedure to generate its edges.

The GVG edges are traced in an incremental manner using an adaptation of numerical continuation techniques [11]. Practically speaking, these techniques trace the roots of an expression $G(x) = 0$. Since the Jacobian of G is surjective [7], the implicit function theorem asserts that the roots of G locally define a GVG edge. A robot can locally construct a GVG edge by numerically tracing the roots of this function.

The GVG is defined in terms of smooth functions, wheras the OPP is not. The curve tracing techniques for smooth functions is adapted to trace the extrema of nonsmooth functions. Using new results in nonsmooth analysis and some new results in convex analysis, it can be shown that the OPP freeways are a subset of the GVG. This is useful because the OPP can now be generated using the GVG edge tracing technique. That is, the same G matrix can generate the OPP freeways (Section 6.1)

3. Distance Function

A function that encodes the distance between a robot and nearby obstacles is key to many motion planning algorithms, such as those mentioned in the previous section. The purpose of this section is to demonstrate the distance function is nonsmooth and is of the form $F(x) = \min_i f_i(x)$, where $F(x)$ is nonsmooth and each $f_i(x)$ is convex.

Assume the robot can be modeled as a point (as it is in configuration spaces) and this point robot is operating in a subset, termed the work space ($\mathcal{W}$), of an m-dimensional Euclidean space. The work space $\mathcal{W}$ is populated by obstacles $C_1, \ldots, C_n$ which are closed convex subsets of $\mathcal{W}$. Non-convex obstacles are modeled as the union of convex shapes. It is assumed that $\mathcal{W}$ is bounded by is a collection of convex sets, which are members of the obstacle set $\{C_i\}$. Finally, the freespace, $\mathcal{FS}$ is the portion of the workspace not occupied by obstacles, i.e.,

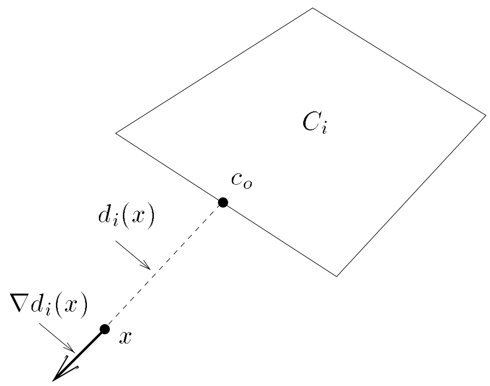

Figure 2: Distance between x and C_i is the distance to the closest point on C_i. The gradient is a unit vector pointing away from the nearest point.

$\mathcal{FS} = \mathcal{W} \setminus \bigcup_i C_i$.

The following are the key distance function definitions:

- single object distance function (Figure 2)

$$d_i(x) = \min_{c \in C_i} \|x - c\|,$$

- single object distance function gradient (Figure 2)

$$\nabla d_i(x) = \frac{x - c_0}{\|x - c_0\|}, \tag{1}$$

  for $c_0 = \operatorname{argmin}_{c \in C_i} \|x - c\|$,

- multi-object distance function (also termed the *distance function*)

$$D(x) = \min_i d_i(x).$$

3.1. Nonsmooth Distance Function

The distance function is nonsmooth. This section motivates this claim by considering the distance function constrained to a slice, which is relevant to the OPP method. A slice is defined as follows. Let $\alpha \colon \mathcal{W} \to \mathbb{R}$ be a smooth function which foliates $\mathcal{W}$, i.e., $\bigcup_{\lambda \in \mathbb{R}} \{x : \alpha(x) = \lambda\} = \mathcal{W}$. A slice Y is the pre-image of λ under α.

For now, assume that a slice is a hyperplane which is defined by the pre-image of a scalar under $\alpha \colon \mathbb{R}^m \to \mathbb{R}$ where $\alpha(x) = x_1$. So, $\alpha(x)$ is the first coordinate of x. In this case, we can decompose the physical space coordinates x into "slice coordinates" y and the "sweep coordinate" λ: $x = [\lambda, y]^T$.

The single object distance function *constrained to a slice* is the distance between a point that is in a slice λ and a set C_i, i.e.,

$$\tilde{d}_i(y; \lambda) = d_i(x) \qquad \text{where } \alpha(x) = \lambda \text{ and } y \in \alpha^{-1}(\lambda). \tag{2}$$

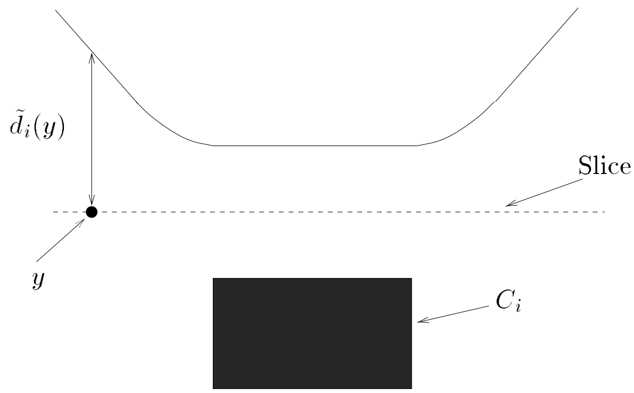

Figure 3: Distance function plotted along a horizontal slice.

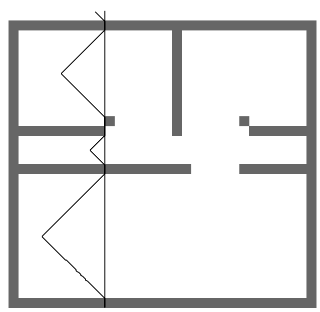

Figure 4: Distance function plotted along a vertical slice. The vertical line is the slice and the graph of the distance is vertically plotted along the slice.

Hereafter, $\tilde{d}_i$ is shorthand for $\tilde{d}_i(y;\lambda)$. See Figure 3 for an example of the distance function plotted along a slice. At each slice point, $\tilde{d}_i$ is computed to the closest point of the obstacle.

Typically, a robot's environment is populated with multiple obstacles, and thus we define a distance function for multiple obstacles. Therefore, there is a multi-object distance function *constrained to a slice*, which measures the distance between a point in a slice λ and the *closest* obstacle to that point, i.e.,

$$\tilde{D}(y;\lambda) = \min_i \tilde{d}_i(y;\lambda). \tag{3}$$

Note that $\tilde{D}$ is not necessarily smooth (at the local maxima), as can be seen in Figures 4 and 5 where $\tilde{D}$ is plotted along a vertical and diagonal slice, respectively.

3.2. Convexity Properties of the Single Object Distance Function

Clarke shows that d_i is Lipschitz, i.e., $|d_i(x) - d_i(y)| \leq \|x - y\|$ (and thus is

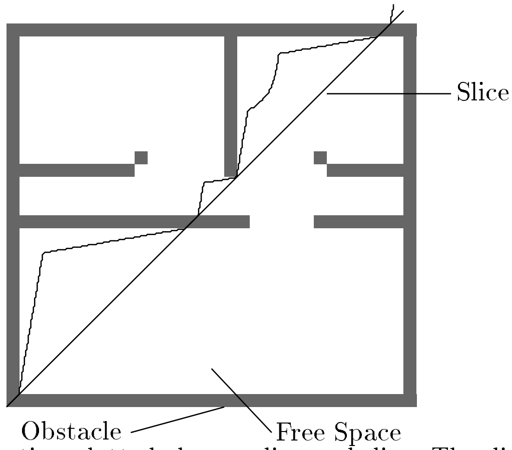

Figure 5: Distance function plotted along a diagonal slice. The diagonal line is the slice and the graph of the distance is diagonally plotted along the slice.

continuous) [8]. He also shows that d_i is convex, but a slightly different proof is included below.

Proposition 1 *If C_i is a convex set, the single object distance function d_i is a convex function.*

Proof. Let $x, y \in \mathcal{FS}$ and $\lambda \in [0,1]$ be given. Pick c_x and c_y in C_i such that

$$\|c_x - x\| = d_i(x), \qquad \|c_y - y\| = d_i(y),$$

and now chose $c \in C_i$ as $c = \lambda c_x + (1-\lambda)c_y$. which must be in C_i because the obstacles are convex. For the same λ, chose $z = \lambda x + (1-\lambda)y$ and pick c_z such that $d_i(z) = \|c_z - z\|$. Without loss of generality, assume $z \notin C_i$, Now, we have

$$\begin{aligned}
d_i(\lambda x + (1-\lambda)y) &= \|c_z - \lambda x - (1-\lambda)y\| && \text{(definition of } d_i\text{)} \\
&\leq \|c - \lambda x - (1-\lambda)y\| && \text{(definition of } d_i\text{)} \\
&= \|\lambda c_x + (1-\lambda)c_y - \lambda x - (1-\lambda)y\| && (c = \lambda c_x + (1-\lambda)c_y) \\
&\leq \|\lambda c_x - \lambda x\| + \|(1-\lambda)c_y - (1-\lambda)y\| && \text{(triangle inequality)} \\
&= \lambda\|c_x - x\| + (1-\lambda)\|c_y - y\| \\
&\leq \lambda d_i(x) + (1-\lambda)d_i(y)
\end{aligned}$$

□

Since d_i is convex, the nonsmooth function $D(x) = \min_i d_i(x)$ is of the form $F(x) = \min_i f_i(x)$, the class of functions considered in this paper.

4. Review of Nonsmooth Analysis

Several useful functions are nonsmooth at many points of interest and thus do not have a conventional derivative at these points. However, one can build a calculus for such nonsmooth functions from a less restrictive class of assumptions than smoothness: Lipschitzness, and convexity. We review here some essential results from nonsmooth analysis [8] and develop a few new useful results.

Recall that a function h is *Lipschitz* at x if for all x and y in a neighborhood of x, $|h(x) - h(y)| \leq K\|x - y\|$, where K is a positive scalar. Lipschitzness is a more restrictive property on functions than continuity, i.e., all Lipschitz functions are continuous. However, all Lipschitz functions are not guaranteed to be smooth. In such cases, they possess a *generalized gradient* [8], i.e.,

Definition 2 (Generalized Gradient) *In a finite dimensional space,* ***the generalized gradient*** *of a Lipschitz function h at a point x is denoted by $\partial h(x)$ and given by:*

$$\partial h(x) = Co\{\lim_{x_i \to x} \nabla h(x_i) : x_i \notin S,\ x_i \notin \Omega_h\},$$

where Ω_h is the set of points where h fails to be differentiable, S is any set of measure zero, and Co means convex hull.

Note that if h is smooth at x, then $\partial h(x)$ reduces to the conventional gradient. The use of the ∂ can be confusing because it can also mean "boundary" when it precedes a variable that represents a set. See the Appendix for the formal definition of the convex hull, which appears in the definition of the generalized gradient.

It can be seen that the scalar multiple of the generalized gradient is the generalized gradient of the scalar multiple [8], i.e., $\partial h(sx) = s\partial h(x)$ for a scalar s. In particular, note that $\partial(-h) = -\partial(h)$.

The maximum of a set of Lipschitz functions is also a Lipschitz function, i.e., $H(x) = \max_i\{h_i(x)\}$ is Lipschitz when all of the $h_i(x)$ mappings are Lipschitz [8]. Therefore, $H(x)$ has a generalized gradient. In fact, if each h_i is convex, the generalized gradient of H is

$$\partial H(x) = \mathrm{Co}\{\partial h_i(x) : i \in I(x)\},$$

where $I(x)$ is the set of indices where $h_i(x) = H(x)$.

Using the definitions and propositions outlined in this section, consider the generalized gradient of $F(x) = \min_i f_i(x)$ where each f_i is convex.

Proposition 2 (Pointwise Minima) *For a set of convex functions $f_i(x)$, the function*

$$F(x) = \min_{i=1,\ldots,n}(f_i(x)) \tag{4}$$

has a generalized gradient given by

$$\partial F(x) = \mathrm{Co}\{\partial f_i(x) : i \in I(x)\}, \tag{5}$$

where $I(x)$ is the set of indices for which $f_i(x) = F(x)$, i.e., where $f_i(x)$ attains the minimum over all $i \in I(x)$.

Proof. Note that $\min_{i=1,\ldots,n} f_i(x) = -\max_{i=1,\ldots,n}(-f_i(x))$ and thus this proof is a simple consequence of Proposition A.1 and $-\partial(f_i(x)) = \partial(-f_i(x))$.

$$\begin{aligned}\partial F(x) &= -\mathrm{Co}\{\partial(-f_i(x)) : i \in I(x)\} \\ &= -\mathrm{Co}\{-\partial(f_i(x)) : i \in I(x)\} \\ &= \mathrm{Co}\{\partial f_i(x) : i \in I(x)\}.\end{aligned}$$

□

5. Extrema of Nonsmooth Functions

From calculus, if x is an extrema of a smooth function, then the smooth function's derivative is zero at x. Clarke shows the nonsmooth analog of this statement [8]: if H attains a local extrema at x, then $0 \in \partial H(x)$. (Note that 0 refers to the origin of the tangent space at $x \in \mathbb{R}^m$.) The converse of this statement, on the other hand, is not always true.

This section derives a converse statement for functions of the form $F(x) = \min_i f_i(x)$. As the following results point out, it is possible to determine extrema from first derivative information. In fact, the results also classify which types of extrema (i.e., maxima, minima, saddle points) are determined, also from first order information. This result bypasses the need to use the Hessian, or other curvature conditions, which classify the extremal points of smooth functions.

Proposition 3 *Let $F\colon \mathbb{R}^m \to \mathbb{R}$ be a mapping where $F(x) = \min_i f_i(x)$ and each f_i is a smooth convex function. For $x^* \in \mathbb{R}^m$, let 0 be the origin of the tangent space $T_{x^*}\mathbb{R}^m$. If $0 \in \mathrm{int}(\partial F(x^*))$ then x^* is a local maximum.*

Proof. This proof relies on the following lemma whose proof appears in the Appendix.

Lemma A.5 *The origin is contained in the interior of the convex hull of a set of n arbitrary vectors $\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\}$ if and only if there exists a v_i such that for all $w \in \mathbb{R}^m$, $\langle w, v_i \rangle > 0$.*

For each direction w, by Lemma A.5, there exists an i for which $\langle w, -\nabla f_i(x^*)\rangle > 0$ because $0 \in \mathrm{Co}\{\nabla f_i(x^*)\}$. In other words, there exists an i for which f_i decreases in the direction of w in a neighborhood of x^*. Since this is true for all w, there is always an f_i that decreases in any direction w. Therefore, there exists a $\delta > 0$ where for all $\delta > \epsilon > 0$,

$$f_i(x + \epsilon w) < f_i(x^*) = F(x^*).$$

By definition, $F(x^* + \epsilon w) \leq f_i(x^* + \epsilon w)$, thus

$$F(x^* + \epsilon w) < F(x^*) \quad \forall w,$$

which implies that $F(x^*)$ is a local maxima. □

Figures 6 and 7 demonstrate the generalized gradient concept for the distance function. The conditions for saddle points are similar and can be proven in a similar fashion.

Proposition 4 *At a saddle point, F is nonsmooth, and the origin is contained in the boundary of $\partial F(x^*)$.*

Finally, for local minima, we have

Proposition 5 *At a local minimum of F, $0 = \partial F(x^*)$.*

Proof. At a local minimum, there must be one unique i for which $F(x) = f_i(x)$. In this case, the generalized gradient reduces to the conventional gradient. At a local minimum, this gradient vanishes. □

6. Applications to Motion Planning

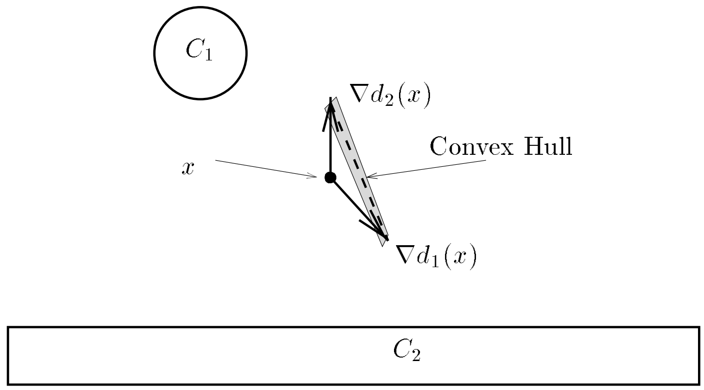

Figure 6: Generalized gradient. Note how the origin is not contained in the interior of the generalized gradient and thus x is not a local maximum. Note that the point x is equidistant to two obstacles.

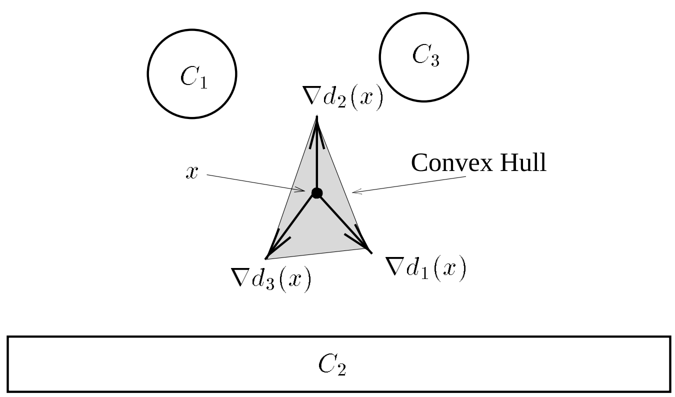

Figure 7: Generalized gradient. Note how the origin is contained in the interior of the generalized gradient and thus x is a local maximum. Note that the point x is equidistant to three obstacles.

This section now applies the results of the previous section and the Appendix to the motion planning problems introduced in Section 2.

6.1. Opportunistic Path Planner

Recall that the OPP freeways are the local maxima of distance constrained to a slice, i.e., the set of points $x^* = [\lambda^*, {y^*}^T]^T$ where $D(y^*; \lambda^*)$ has its maximum on a slice (and not necessarily in the ambient space). For the OPP method, the slice Y is a co-dimension one space that is isometric to $\mathbb{R}^{m-1}$, i.e., it is a hyperplane. This section derives the local maxima for $F = \min f_i(x)$ but restricted to a slice parameterized by λ, i.e.,

$$\tilde{f}_i(y; \lambda) = f_i(x) \qquad \text{where } \alpha(x) = \lambda \text{ and } y \in \alpha^{-1}(\lambda). \tag{6}$$

With this definition in hand, consider the following function:

$$\tilde{F}(y; \lambda) = \min_i \tilde{f}_i(y; \lambda). \tag{7}$$

The generalized gradient of F restricted to a slice is described by the following proposition. For the following proposition, note that π_Y orthogonally projects vectors onto the Y subspace and ∂_Y represents partial differentiation with respect to all variables in the Y subspace.

Proposition 6 *The orthogonal projection of $\partial F(x)$ onto the Y subspace is equal to the partial gradient of $\tilde{F}(y)$ with respect to Y. In other words,*

$$\pi_Y(\partial F(x)) = \partial_Y \tilde{F}(y; \lambda).$$

Proof. First recall that for smooth functions that the partial derivative is equal to projection of the full derivative, i.e.,

$$\nabla_{X_1} f(x_1, x_2) = \pi_{X_1} \left(\nabla f(x_1, x_2)\right),$$

where $f\colon X_1 \times X_2 \to \mathbb{R}$ is a differentiable function. If there is a unique i for which $F(x) = f_i(x)$, then $F(x)$ is smooth at x, and the proposition is proved. If there is not a unique i, by Proposition 2,

$$\partial F(x) = \sum_{i \in I(x)} \eta_i \nabla d_i(x) \quad \text{such that} \sum_{i \in I(x)} \eta_i = 1 \text{ and } \eta_i > 0$$

Now project this generalized gradient onto the Y coordinates:

$$\begin{aligned} \pi_Y(\partial F(x)) &= \pi_Y\left(\textstyle\sum_{i \in I(x)} \eta_i \nabla f_i(x)\right) \\ &= \textstyle\sum_{i \in I(x)} \pi_Y(\eta_i \nabla f_i(x)) \\ &= \textstyle\sum_{i \in I(x)} \eta_i \pi_Y \nabla f_i(x) \\ &= \textstyle\sum_{i \in I(x)} \eta_i \nabla_Y(\tilde{f}_i(y; \lambda)) \\ &= \partial_Y \tilde{F}(y). \end{aligned}$$

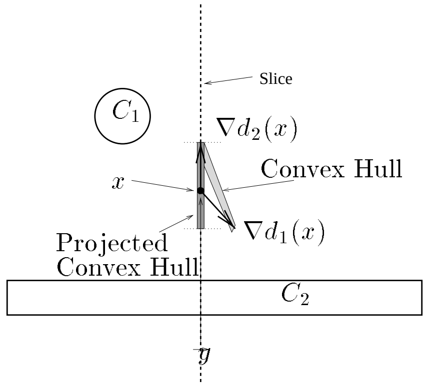

Figure 8: Generalized gradient. Note that the origin is contained in the interior of the projected generalized gradient and thus x is a local maximum.

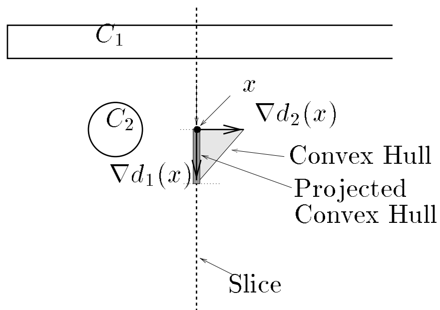

Figure 9: Saddle point because 0 is contained in the boundary of $\pi_Y \partial D(x)$.

$\square$

Propositions 2,3, and 6 assert the following corollary, which yields the first derivative test for nonsmooth functions constrained to a slice, thereby allowing the definition of the OPP freeways.

Corollary 1 Let $x^* = (\lambda, y^*)$, $Y = \alpha^{-1}(\lambda)$, and $T_Y\mathbb{R}^{m-1} = \pi_y T_{x^*}\mathbb{R}^m$ be a co-dimension one slice (isometric to $\mathbb{R}^{m-1}$).

$$0 \in \mathrm{int}(\mathrm{Co}(\{\nabla_Y \tilde{f}_i(y^*;\lambda)\})) \iff 0 \in \partial_Y \tilde{F}(y^*;\lambda),$$

and if $0 \in \mathrm{int}(\mathrm{Co}(\{\nabla \tilde{f}_i(y^*;\lambda)\}))$, then y^* is a local maxima of $\tilde{F}$ on the co-dimension slice.

A similar result applies to saddle points and minima constrained to a slice. Since $\tilde{D}$ is member of the class of functions $\tilde{F}$, Corollary 1 supplies the local maxima of the distance function constrained to a slice and Figures 8, 9, and 10 demonstrate these concepts.

Finally, it is interesting to note the following proposition whose proof is omitted due to space limitations

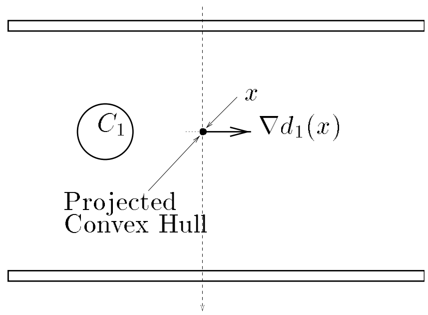

Figure 10: Local minimum because $0 = \pi_Y \partial D(x)$.

Proposition 7 *If $x \in \mathbb{R}^m$ is a local maximum of F in the ambient space, then x is a local maximum of F in all slices which pass through x.*

6.2. Accessibility of the Generalized Voronoi Graph

Recall that the GVG is the locus of points equidistant to m obstacles in m dimensions such the m obstacles are the closest obstacles, i.e.,

$$\{x \in \mathcal{FS} : 0 \leq d_{i_1}(x) = \cdots = d_{i_m}(x) \leq d_h(x) \;\; \forall h\}.$$

Accessibility is the property that a path can be constructed from any point in the free space to a point on the GVG. A sequence of gradient ascent operations achieves accessibility for the GVG. The following proposition supplies an argument that describes a path from any point in the freespace to a GVG edge.

Proposition 8 *In a bounded environment, the GVG has the property of accessibility.*

Proof. We demonstrate that a robot can access the GVG by following a path that is constructed using gradient ascent on the multi-object distance function $D(x)$, which is the distance to the nearest object from x. It has already been shown that D is nonsmooth and thus exhibits a generalized gradient. Furthermore, it is shown in Section 5 that if $0 \in \text{int}(\partial D(x))$, where 0 is the origin of the tangent space at x, then x is a local maxima of D. Using this result and the following two lemmas whose proofs appear in the appendix, we can conclude that if x is a local maxima of D, then the point x is equidistant to $m+1$ obstacles.

Lemma A.2 [Goldman and Tucker] *It requires a minimum of $(m+1)$ vectors to positively span $\mathbb{R}^m$.*

Lemma A.4 *Given a set of n arbitrary vectors in $\mathbb{R}^m$, then $0 \in \text{int}(\text{Co}\{v_i \in \mathbb{R}^m : \; i = 1, \ldots, n\})$ if and only if $\{v_i \in \mathbb{R}^m : \; i = 1, \ldots, n\}$ positively span $\mathbb{R}^m$.*

The results of Scheimber and Olivera [12] state that the second derivative of D, termed *a generalized Hessian*, is always positive or contains only positive values. In a sense, the graph of the function D is always "concave up." Therefore, the generalized gradient of D only vanishes at a local minima. Assume the robot does not start at a local minima, which is reasonable because since we are performing a gradient ascent operation, the local minima are unstable extrema points. That is, if the robot were slightly perturbed from a local minima, it will escape. Since the local minima occur on a set of measure zero, they can be practically ignored. Therefore, gradient ascent of the multi-object distance function will bring the robot to a local maxima of D, which is a point equidistant to $m+1$ obstacles and thereby is a point on the GVG. (Note that when ∂D is a set, the vector with the smallest norm in ∂D is chosen as the gradient [12].) □

6.3. Incremental Construction of the OPP (OPP Freeways are a Subset of the GVG)

The OPP roadmap does not have an incremental construction procedure, whereas the GVG does. The GVG edges are traced in an incremental manner using an adaptation of numerical continuation techniques [11]. Let x be the coordinates of a point on a GVG edge. At $x \in \mathbb{R}^m$, assign a local coordinate system (y, λ) such that λ points along the tangent of the GVG edge and the y coordinates spans Y, the hyperplane orthogonal to the GVG edge. Let Y be termed the "normal plane."

Assume the robot starts at a point x on the GVG (i.e., assume it just accessed the GVG). Numerical techniques trace the roots of an expression $G(y, \lambda) = 0$ as the parameter λ varies, where

$$G(y,\lambda) = \begin{bmatrix} d_1(y,\lambda) - d_2(y,\lambda) \\ \vdots \\ d_1(y,\lambda) - d_m(y,\lambda) \end{bmatrix} \tag{8}$$

where d_i is the single object distance function to the m closest obstacles. Since G is a function of distance, it can be computed from sensor readings.

Note that the function $G(y, \lambda)$ assumes a zero value only on the GVG. Hence, if the Jacobian of G is surjective [7], then the implicit function theorem implies that the roots of $G(y, \lambda)$ locally define a GVG edge as λ is varied. A robot can locally construct a GVG edge by numerically tracing the roots of this function.

Fortunately, the OPP approach can borrow the GVG incremental construction procedure for sensor based implementation of OPP. Using the new results derived in this paper for nonsmooth analysis and convex analysis, it is shown below that the OPP freeways are a subset of the GVG. This is useful because the OPP can now be generated using the GVG edge tracing technique. That is, the same G function can generate the OPP freeways.

Proposition 9 *The freeways of the OPP method are a subset of the GVG edges.*

Proof. Proposition 3 states that at a point $x \in \mathbb{R}^m$, if zero is in the interior of the generalized gradient of the multi-object distance function, D, then D attains a local maximum at x. Furthermore Corollary 1 asserts that if zero is in the projection of the interior of the generalized gradient, then D, constrained to the slice, attains a local maximum at x. That is, for $x^* = (\lambda^*, {y^*}^T)^T \in \mathbb{R}^m$, if $0 \in \text{int}(\partial_Y D(y^*; \lambda)) = \text{int}(\pi_Y \partial D(y^*; \lambda))$, then y^* is a local maximum on a slice.

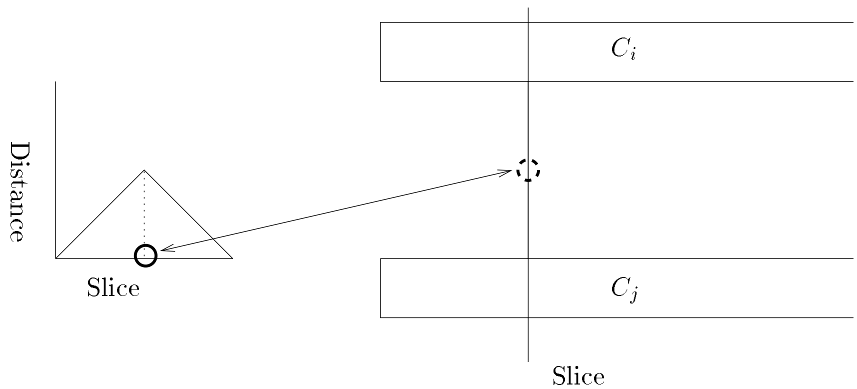

Figure 11: The local maxima of the multi-object distance function occurs only at equidistance. The two circles both correspond to the same point. The dashed circle surrounds the point, equidistant to objects C_i and C_j. The solid circle surrounds the point on the slice where distance is maximized, as depicted by the graph of the distance function in the left-hand side.

In the OPP method, the freeways are the trace of local maxima of the multi-object distance function constrainted to a slice. Proposition 3 indicates that for all points on the freeway, the origin must be in the generalized gradient of the multi-object distance function, ∂D, constrained to a slice. Recall that x is equidistant to h obstacles, if and only if the generalized gradient of D is the convex hull of h single object distance gradient vectors.

By Lemma A.4, the origin is in the convex hull of a collection of vectors if and only if those vectors positively span the space. Therefore, by Lemma A.4 and Proposition 3, if x is a local maximum of D, constrained to a slice, then the single object distance gradient vectors positively span $T_x\mathbb{R}^{m-1}$ which is diffeomorphic to $\mathbb{R}^{m-1}$. Lemma A.2 guarantees that the generalized gradient of D is the convex hull of at least m gradient vectors. Therefore, x is equidistant to at least m obstacles at a local maxima of D.

When x is a local maxima of D, not only is it equidistant to m objects, but by definition of D, the point x is closer to these m objects than to any other object. Thus, the OPP freeway segments are a subset of the GVG edges. $\square$

Figure 11 depicts the correspondence between local maxima of D and equidistance of points in the GVG to nearby obstacles.

However, it also needs to be shown that the Jacobian of G is invertible for the numerical procedures to work. The following proposition guarantees that ∇G is invertible in a neighborhood of the OPP freeway.

Proposition 10 *In a neighborhood of a point x on a freeway, $\nabla_Y G(x)$ is invertible.*

Proof. It will be shown that ∇G is invertible on a freeway, and then by continuity of the distance function and the determinant function, $\nabla G(x)$ is invertible in a neighborhood of the OPP freeways. This proof relies on the following lemma whose proof appears in the Appendix.

Lemma A.6 *If $\{v_1, \ldots, v_m\}$ postively span $\mathbb{R}^{m-1}$, then $\{v_1 - v_2, v_1 - v_3, \ldots, v_1 - v_m\}$ span $\mathbb{R}^{m-1}$.*

Lemma 1 *At a point x on a freeway, $\nabla_Y G(x)$ is invertible.*

Proof. Let $x = (\lambda, y^T)^T$ be the coordinates where the y coordinates span the hyperplane, Y, orthogonal to the sweep direction and λ corresponds to the sweep direction. Proposition 3 states that if $0 \in \mathrm{int}(\partial_Y D(y; \lambda)) = \mathrm{int}(\pi_Y \partial D(y; \lambda))$, then y is a local maximum. Therefore, if $x = (\lambda, y^T)^T$ is a point on a freeway, then $0 \in \mathrm{Co}\{\nabla_Y d_1, \ldots, \nabla_Y d_m\}$.

By Lemma A.4, $0 \in \mathrm{Co}\{\nabla_Y d_1, \ldots, \nabla_Y d_m\}$ if and only if the vectors $\{\nabla_Y d_1, \ldots, \nabla_Y d_m\}$ positively span the slice. Therefore, at a local maxima (i.e., a point on a freeway), $\{\nabla_Y d_1, \ldots, \nabla_Y d_m\}$ positively span the slice.

Finally, by Lemma A.6, since $\{\nabla_Y d_1, \ldots, \nabla_Y d_m\}$ positively span the slice, $\{\nabla_Y d_1 - \nabla_Y d_2, \nabla_Y d_1 - \nabla_Y d_3, \ldots, \nabla_Y d_1 - \nabla d_m\}$ span the slice. That is, they are linearly independent.

Therefore,

$$\nabla_Y G(y, \lambda) = \begin{bmatrix} (\nabla_Y d_1(y,\lambda) - \nabla_Y d_2(y,\lambda))^T \\ (\nabla_Y d_1(y,\lambda) - \nabla_Y d_3(y,\lambda))^T \\ \vdots \\ (\nabla_Y d_1(y,\lambda) - \nabla_Y d_m(y,\lambda))^T \end{bmatrix}$$

is invertible. ∇

Since the determinant function and distance function are continuous, there exists a neighborhood about the OPP freeways for which $\nabla_Y G$ is invertible. $\square$

Since the OPP is a subset of the GVG and the Jacobian of G is invertible near the OPP freeways, the same continuation method that generates the GVG also incrementally constructs the OPP roadmap.

7. Conclusion

A bulk of this paper is devoted to nonsmooth analysis of a class of functions $F(x) = \min_i f_i(x)$ where $f_i(x)$ is a smooth convex function. Using only first order information, a methodological approach was developed to determine local maxima, local minima, and local saddle points of this class of nonsmooth functions. Since motion planning originally motivated this work, the distance function, which provides the basis of many motion planners, was rigorously analyzed.

The results of this paper were applied to Canny and Lin's OPP method which generates a roadmap by looking at local maxima of the distance function D, which is a nonmsooth function. The nonsmooth analysis gives a condition for when D is locally maximized. Further analysis, which exploits the nonsmooth properties of D, provided an incremental construction procedure for tracing out the edges of the OPP roadmap. This incremental construction procedure allows for the generation of OPP roadmap edges (and GVG edges) in an arbitrary m-dimensional space. The nonsmooth analysis of D was also applied to Choset and Burdick's generalized Voronoi graph to prove the accessibility criterion.

Although this paper's applications focused on the nonsmooth analysis of D as it applies to Canny and Lin's Opportunistic Path Planner and Choset and Burdick's generalized Voronoi graph, the analysis of D applies to all motion planning algorithms, such as potential functions that use distance in its algorithms. Further, this analysis is good for sensor based planning because many sensor based planners use distance information because sensors readily provide such information. The nonsmooth and convex analysis results are not limited to motion planning.

Finally, many of the nonsmooth analysis results here depend upon some results in convex analysis which are listed and proven in the appendix. Some of the lemmas in the appendix have been only partially proven in previous literature.

Acknowledgements

The author would like to gratefully thank Teresa McMullen at ONR, Grant 97PR06977 and Howard Moraff and Larry Reeker at NSF, Grant IRI-9702768 for supporting this work. The author acknowledges his Ph.D. advisor Joel Burdick for his input and direction in this work. Also, Jim Ostrowski, Bill Messner, and Andrew Lewis were helpful in deriving some of the proofs in this work. Also, many thanks go to Jon Luntz for thoroughly proofreading this paper.

References

1. J.F. Canny, *The Complexity of Robot Motion Planning,* (MIT Press, Cambridge, MA, 1988).
2. J.C. Latombe, *Robot Motion Planning,* (Kluwer Academic Publishers, Boston, MA, 1991).
3. J. E., Goodman., J. O'Rourke, *Handbook of Discrete and Computational Geometry* (CRC Press, New York, 1997).
4. J.F. Canny, M.C. Lin, "An Opportunistic Global Path Planner," *Algorithmica* **Vol 10.** (1993) (102–120)
5. C. Ó'Dúnlaing and C.K. Yap, "A "Retraction" Method for Planning the Motion of a Disc," *Algorithmica*, **6** (1985) 104–111.
6. H. Choset, J. Burdick, "Sensor Based Planning, Part I: The Generalized Voronoi Graph" *Proc. IEEE Int. Conf. on Robotics and Automation*, Nagoya, Japan, 1995.
7. H. Choset, J. Burdick, "Sensor Based Planning, Part II: Incremental Construction of the Generalized Voronoi Graph" *Proc. IEEE Int. Conf. on Robotics and Automation*, Nagoya, Japan, 1995.
8. F. H. Clarke, *Optimization and Nonsmooth Analysis* (Society of Industrial and Applied Mathematics, Philadelphia, PA, 1990).
9. A. J. Goldman, A. W. Tucker, "Linear Inequalities and Related Systems," in *Polyhedral Convex Cones*, (Princeton University Press, Princeton, NJ, 1956).
10. O. Khatib, "Real-Time Obstacle Avoidance for Manipulators and Mobile Robots," in *International Journal of Robotics Research*, **5** (1995) 90–98.
11. H.B. Keller, *Lectures on Numerical Methods in Bifurcation Problems,* (Tata Institute of Fundamental Research, Bombay, India, 1987).
12. S. Scheimberg, P.R. Oliveira, "Descent Algorithm for a Class of Convex Nondifferentiable Functions," *Journal of Optimization Theory and Applications,* **72** (February,

1992) 269–297.

13. R. M. Murray, Z. Li, S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation,* (CRC Press, Boca Raton, FL, 1994).
14. V-D. Nguyen, "Constructing Force-closure Grasps," *International Journal of Robotics Research,* **7** (June, 1988) (3–16).

Appendix A: Convex Analysis

Convex analysis [9] is the study of functions that are convex. It has been widely used in robotic grasping [13]. Two of the following results have been stated without proof [13,14]. Although the results of this section support this work in nonsmooth analysis, these results are independent of it.

Definition A.1 (Convex Function) *The function* $f\colon \mathbb{R}^m \to \mathbb{R}$ *is* ***convex*** *if for all* $x_i, i = 1, \dots, n$, $\sum_{i=1}^{n} \lambda_i = 1$, *and* $0 \le \lambda_i \le 1$,

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \le \sum_{i=1}^{n} \lambda_i f(x_i).$$

Definition A.2 (Convex Hull) *The* ***convex hull*** *of the set of vectors* $\{v_i : i = 1, \dots, n\}$ *is*

$$\mathrm{Co}\{v_i : i = 1, \dots, n\} = \left\{\sum_{i=1}^{n} \lambda_i v_i : \lambda_i \in \mathbb{R} \textit{ such that } \lambda_i > 0 \ \forall i \ \textit{ and } \sum_{i=1}^{n} \lambda_i = 1\right\}.$$

Proposition A.1 The negated convex hull of a set of vectors is the convex hull of the negated vectors, i.e.,

$$-\mathrm{Co}\{x_1, \dots, x_n\} = \mathrm{Co}\{-x_1, \dots, -x_n\}. \tag{A.1}$$

Proof.

$$\begin{aligned} -\mathrm{Co}\{x_1, ..., x_n\} &= -\{\textstyle\sum_{i=1}^{n} \lambda_i x_i\} \\ &= \{\textstyle\sum_{i=1}^{n} -\lambda_i x_i\} \\ &= \{\textstyle\sum_{i=1}^{n} \lambda_i(-x_i)\} \\ &= \mathrm{Co}\{-x_1, \cdots, -x_n\} \end{aligned}$$

□

The following lemmas will show that there is a relationship between vectors which span a space, positively span a space, and whose convex hull contains the origin of a space. This section enumerates some lemmas that were only partially proven in the past.

Lemma A.1 (Nguyen and Goldman & Tucker) *In an* m*-dimensional vector space, a set of vectors* $\{v_i \in \mathbb{R}^m : i = 1, \dots, n\}$ *has at least* $m+1$ *vectors* $\{v_i \in \mathbb{R}^m : i = 1, \dots, m+1\}$ *which positively span* $\mathbb{R}^m$ *if and only if*

1. m *of the* $m+1$ *vectors are linearly independent, and*
2. *a strictly positive sum of the* $m+1$ *vectors is the zero vector, i.e.,* $\sum_{i=1}^{i=m+1} a_i v_i = 0$ *where* $a_i > 0$ *for all* i.

The above Lemma is stated without proof in prior work [14] and the proof of the converse also appears in prior literature [9].

Proof. Assume without loss of generality that $n = m + 1$. First, assume that the vectors $\{v_i \in \mathbb{R}^m : i = 1, \ldots, m+1\}$ positively span $\mathbb{R}^m$, which asserts that for all $w \in \mathbb{R}^m$,

$$w = \sum_{i=1}^{m+1} a_i v_i$$

where $a_i > 0$. This implies that for 0, there exists a positive set of integers b_i, such that

$$0 = \sum_{i=1}^{m+1} b_i v_i.$$

Let $c_i = \frac{a_{m+1}}{b_{m+1}} b_i$. Therefore, $0 = \sum_{i=1}^{m+1} c_i v_i$, and thus

$$w - 0 = \sum_{i=1}^{m+1} (a_i - c_i) v_i = \sum_{i=1}^{m} d_i v_i.$$

Note that $d_{m+1} = 0$ because $a_{m+1} - c_{m+1} = a_{m+1} - \frac{a_{m+1}}{b_{m+1}} b_{m+1} = 0$. Since we can represent any arbitrary $w \in \mathbb{R}^m$ with $\{v_i : i = 1, \ldots, m\}$, m of the original vectors spans $\mathbb{R}^m$.

The converse follows similarly. The vectors $\{v_i : i = 1, \ldots, m\}$ span $\mathbb{R}^m$ if and only if they span $\mathbb{R}^m$ translated by v_{m+1}, which is in turn $\mathbb{R}^m$. Therefore, for all $w \in \mathbb{R}^m$,

$$w = \sum_{i=1}^{m} d_i v_i + v_{m+1}.$$

Also, note that there exists a set of positive integers, c_i, such that $0 = \sum_{i=1}^{m+1} c_i v_i$. Let $b_i = \frac{(1+\epsilon)|\min_i d_i|}{\min c_i} c_i$, for a small $\epsilon > 0$. Hence, $0 = \sum_{i=1}^{m+1} b_i v_i$. From this, we have

$$w + 0 = \sum_{i=1}^{m} (d_i + b_i) v_i + (1 + b_{m+1}) v_{m+1}.$$

Let $a_i = d_i + b_i$ for $i = 1, \ldots, m$. It can be seen that $a_i > 0$ because

$$\begin{aligned} d_i + b_i &= d_i + \tfrac{(1+\epsilon)|\min_i d_i|}{\min_i c_i} c_i \\ &= d_i + \tfrac{(1+\epsilon) c_i}{\min c_i} |\min_i d_i| \\ &\geq d_i + (1+\epsilon)|\min_i d_i| \\ &> 0. \end{aligned}$$

Finally, let $a_{m+1} = 1 + b_{m+1}$ which is always positive. Since any w can be written as the positive sum of $\{v_i : i = 1, \ldots, m+1\}$, they positively span $\mathbb{R}^{m+1}$. □

Lemma A.2 (Goldman and Tucker) *It requires a minimum of $(m+1)$ vectors to positively span $\mathbb{R}^m$.*

Proof. Assume that the vectors $\{v_1, \ldots, v_m\}$ span $\mathbb{R}^m$ (a minimum of m vectors is required to span $\mathbb{R}^m$). For all $w \in \mathbb{R}^m$, $w = \sum_{i=1}^{m} a_i v_i$ where $a_i \in \mathbb{R}$.

Define the vector, v_{m+1}, which is the negated sum of the other v_i: $v_{m+1} = -v_1 - v_2 - \cdots - v_m$. Let $\kappa = (\max_i |a_i|) + \epsilon$ where $\epsilon > 0$. Then the following holds for all $w \in \mathbb{R}^m$,

$$\begin{aligned} w &= \textstyle\sum_{i=1}^{m} a_i v_i \\ &= \textstyle\sum_{i=1}^{m} (a_i + \kappa) v_i + \kappa v_{m+1} \\ &= \textstyle\sum_{i=1}^{m+1} b_i v_i. \end{aligned} \tag{A.2}$$

where for $i = 1, \ldots, m$, $b_i = a_i + \kappa$ and $b_{m+1} = \kappa$. Since $\kappa > 0$ and $\kappa > |a_i|$ for all i, all the b_i's are positive. Therefore, $m+1$ vectors may positively span $\mathbb{R}^m$.

Now, it needs to be shown that a minimum of $m+1$ vectors are required to positively span $\mathbb{R}^m$. Recall $v_{m+1} = -v_1 - v_2 - \cdots - v_m$. Clearly, v_{m+1} is an element of $\mathbb{R}^m$ and it is not positively spanned by the other m vectors $\{v_1, \ldots, v_m\}$. □

Lemma A.3 (Messner) *A set of n arbitrary vectors $\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\}$ positively spans $\mathbb{R}^m$ if and only if there exists a v_i such that for all $w \in \mathbb{R}^m$, $\langle w, v_i \rangle > 0$.*

Proof. First consider the "only if" by assuming that $\{v_i\}$ positively span a space. Let $w = \sum_{i=1}^{n} \alpha_i v_i$ be a nonzero vector with $\alpha_i > 0$.

$$\begin{aligned} 0 &< \langle w, w \rangle \\ &= \langle \textstyle\sum_{i=1}^{n} \alpha_i v_i, w \rangle \\ &= \textstyle\sum_{i=1}^{n} \langle \alpha_i v_i, w \rangle. \end{aligned} \tag{A.3}$$

This implies that there exists an i where $\langle \alpha_i v_i, w \rangle > 0$ and since $\alpha_i > 0$, we can conclude that $\langle v_i, w \rangle > 0$ for at least one i.

Now consider the "if" by assuming that for all $w \in \mathbb{R}^m$, there exits an i such that $\langle v_i, w \rangle > 0$. The approach to this proof is to make successive approximations of w by adding positive multiples of the vectors in $\{v_i\}$. Without loss of generality, assume $\|v_i\| = 1$ for all i.

Define for use later

$$\alpha = \inf_{\|y\|=1} \max_i \langle y, v_i \rangle < 1.$$

Note that since the unit ball is compact and $\max_i \langle y, v_i \rangle$ is a continuous function, the infimum is achieved over $\|y\| = 1$ for a specific y_{min}. Therefore, $\alpha > 0$.

Now, we will define the some sequences $\{w_k\}$, $\{e_k\}$, $\{M_k\}$, $\{l_k\}$ and $\{\gamma_{jk}\}$ for $j = 1, \ldots, n$. Let

$$w_0 = 0 \qquad e_0 = w. \tag{A.4}$$

Also let M_k and l_k be be defined such that

$$M_k = \max_j \langle e_k, v_j \rangle \tag{A.5}$$

and l_k is the smallest integer l such that $\langle e_k, v_l \rangle = M_k$. Note that $M_k > 0$. Now define

$$\gamma_{jk} = \sum_{l_m=j,m=0}^{m=k} M_m \tag{A.6}$$

Note that $\{\gamma_{jk}\}$ is a non decreasing sequence since $M_m > 0$. Now define

$$e_{k+1} = e_k - M_k v_{l_k} \tag{A.7}$$

Now, after substituting in for the recursive definition of e_k, we get

$$\begin{aligned} e_{k+1} &= w - \textstyle\sum_{m=0}^{k} M_m v_{l_m} \\ &= w - \textstyle\sum_{j=1}^{n} \left(\sum_{l_m=j,m=0}^{m=k} M_m \right) v_j \\ &= w - \textstyle\sum_{j=1}^{n} \gamma_{jk} v_j \end{aligned} \tag{A.8}$$

Define

$$w_{k+1} = \sum_{j=1}^{n} \gamma_{jk} v_j \tag{A.9}$$

Thus, by definitions of $\{e_k\}$ and $\{\gamma_{jk}\}$, $e_k = w - w_k$.

Now, examine the properties of these sequences. The goal is to show that $e_k \to 0$ as $k \to \infty$ implying that $w_k \to w$ and $\gamma_{jk} \to \bar{\gamma}_j$ as $k \to \infty$ and therefore, $w = \sum_{j=1}^{n} \bar{\gamma}_j v_j$ where $\bar{\gamma}_j \geq 0$.

First, note that by the properties of inner products

$$\langle e_k, v_i \rangle = \|e_k\| \langle \frac{e_k}{\|e_k\|}, v_i \rangle \leq \|e_k\|, \tag{A.10}$$

because $\|v_i\| = 1$. Therefore by definition, $M_k \leq \|e_k\|$.

Also, note that since α is the infimum,

$$M_k = \max_j \langle e_k, v_j \rangle = \max_j \|e_k\| \langle \frac{e_k}{\|e_k\|}, v_j \rangle \geq \alpha \|e_k\| \tag{A.11}$$

Now, examine e_{k+1}

$$\begin{aligned} \|e_{k+1}\|^2 &= \langle e_{k+1}, e_{k+1} \rangle \\ &= \langle e_k - M_k v_{l_k}, e_k - M_k v_{l_k} \rangle \\ &= \langle e_k, e_k \rangle - 2M_k \underbrace{\langle e_k, v_{l_k} \rangle}_{M_k} + \underbrace{\langle v_{l_k}, v_{l_k} \rangle}_{1} M_k^2 \\ &= \|e_k\|^2 - M_k^2 \\ &\leq \|e_k\|^2 - \alpha^2 \|e_k\|^2 \\ &= (1 - \alpha^2) \|e_k\|^2 \end{aligned} \tag{A.12}$$

Therefore, $\|e_k\| \leq (1-\alpha^2)^{\frac{k}{2}} \|e_0\|$, and thus $\lim_{k\to\infty} e_k = 0$. Likewise, $\lim_{k\to\infty} w_k = w$.

Now, consider the coefficients. Recall that $M_k \leq \|e_k\| \leq (1-\alpha^2)^{\frac{k}{2}}\|e_0\|$.

$$\begin{array}{ll}\gamma_{jk} & = \sum_{l_m=j,m=0}^{m=k} M_m \\ & \leq \sum_{l_m=j,m=0}^{m=k} \|e_m\| \\ & \leq \sum_{m=0}^{m=k} \|e_m\| \\ & \leq \sum_{m=0}^{k} (1-\alpha^2)^{\frac{m}{2}} \|e_0\| \\ & \leq \|e_0\| \sum_{lm=0}^{\infty} (1-\alpha^2)^{\frac{m}{2}} \\ & = \|e_0\| \frac{1}{1-(1-\alpha^2)^{\frac{1}{2}}}\end{array} \tag{A.13}$$

Since $\{\gamma_{jk}\}$ is a non-negative non-decreasing sequence that is bounded above, it has a limit $\bar{\gamma}_j \geq 0$.

Therefore,

$$\begin{array}{ll}w & = \lim_{k\to\infty} w_k \\ & = \lim_{k\to\infty} \sum_{j=1}^{n} \gamma_{jk} v_j \\ & = \sum_{j=1}^{n} \bar{\gamma}_j v_j\end{array} \tag{A.14}$$

Since w is arbitrary, $\{v_i\}$ positively span $\mathbb{R}^m$. □

Lemma A.4 *Given a set of n arbitrary vectors in $\mathbb{R}^m$, then $0 \in \mathrm{int}(\mathrm{Co}\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\})$ if and only if $\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\}$ positively span $\mathbb{R}^m$.*

Proof. Assume that $0 \in \mathrm{int}(\mathrm{Co}\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\})$. Let $w \in \mathbb{R}^m$ and chose a positive $s \in \mathbb{R}$ such that $w_s = sw \in \mathrm{nbhd}(0) \subset \mathrm{int}(\mathrm{Co}\{v_i\})$. Therefore,

$$w_s = \sum_i \bar{b}_i v_i \qquad \text{where } \bar{b}_i > 0 \text{ and } \textstyle\sum_i \bar{b}_i < 1.$$

Let $b_i = \frac{\bar{b}_i}{s}$. Therefore,

$$w = \sum_i b_i v_i \qquad \text{where } b_i > 0,$$

and thus since w is arbitrary, $\{v_i\}$ positively span $\mathbb{R}^m$.

Let $w \in \mathbb{R}^m$ be written as

$$w = \sum_i a_i v_i$$

where $a_i \in \mathbb{R}$. Note that some a_i's can be zero.

Assume that $\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\}$ positively span $\mathbb{R}^m$. By Lemma A.1, there exists a set of $\{b_i \in \mathbb{R}\}$ such that

$$0 = \sum_i b_i v_i \qquad \text{where } b_i > 0.$$

Now, let $b = \sum b_i$ and $\bar{b_i} = \frac{b_i}{b+\epsilon}$ where ϵ is a small number greater than zero. Therefore, we have

$$0 = \sum_i \bar{b}_i v_i \qquad \text{where } b_i > 0 \text{ and } \textstyle\sum_i \bar{b}_i < 1.$$

Therefore, 0 is in the interior of the convex hull of $\{v_i\}$. □

And as a simple consequence of Lemmas A.3 and A.4, we have

Lemma A.5 *The origin is contained in the interior of the convex hull of a set of n arbitrary vectors $\{v_i \in \mathbb{R}^m : i = 1, \ldots, n\}$ if and only if there exists a v_i such that for all $w \in \mathbb{R}^m$, $\langle w, v_i \rangle > 0$.*

Lemma A.6 *If $\{v_1, \ldots, v_m\}$ postively span $\mathbb{R}^{m-1}$, then $\{v_1 - v_2, v_1 - v_3, \ldots, v_1 - v_m\}$ span $\mathbb{R}^{m-1}$.*

Proof. Since $\{v_1, \ldots, v_m\}$ postively span $\mathbb{R}^{m-1}$, there exists a set of nonzero $\{a_i \in \mathbb{R}\}$ such that $0 = \sum_{i=1}^m a_i v_i$ (Lemma A.1). In fact, by Lemma A.1, we can write one of the vectors as a linear combination of the others. Without loss of generality, $v_1 = -\sum_{i=2}^m \frac{a_i}{a_1} v_i$. For $i > 2$, let $b_i = \frac{a_i}{a_1}$. So, $v_1 = -\sum_{i=2}^m b_i v_i$.

$$\begin{aligned}
\textstyle\sum_{i=2}^m c_i(v_i - v_1) = 0 \quad &\iff \textstyle\sum_{i=2}^m c_i v_i - \sum_{i=2}^m c_i v_1 = 0 \\
&\iff \textstyle\sum_{i=2}^m c_i v_i - v_1 \sum_{i=2}^m c_i = 0 \\
&\iff \textstyle\sum_{i=2}^m c_i v_i + \left(\sum_{i=2}^m b_i v_i\right)\left(\sum_{i=2}^m c_i\right) = 0 \\
&\iff \textstyle\sum_{i=2}^m \left(c_i + \left(\sum_{i=2}^m c_i\right) b_i\right) v_i = 0 \\
&\iff \textstyle\left(\left(\sum_{i=2}^m c_i\right)\left(1 + \left(\sum_{i=2}^m b_i\right)\right)\right) v_i = 0 \\
&\iff \textstyle\sum_{i=2}^m c_i = 0 \quad \text{or} \quad \sum_{i=2}^m b_i = -1
\end{aligned} \tag{A.15}$$

$\sum_{i=2}^m b_i = -1$ cannot be true because $\forall i$, $b_i > 0$, Since by Lemma A.1 the set $\{v_2, \ldots v_m\}$ spans $\mathbb{R}^{m-1}$, $\sum_{i=2}^m c_i v_i = 0$ if and only if $c_i = 0$ for all i. Therefore, $\sum_{i=2}^m c_i(v_i - v_1) = 0$ if and only if $c_i = 0$ for all i and thus $\{v_i - v_1 : i = 2, \ldots, m\}$ are linearly independent. These $m - 1$ linearly independent vectors span $\mathbb{R}^{m-1}$. □