UBC CPSC 536N: Sparse Approximations Winter 2013

Lecture 3 — January 9, 2013

Prof. Nick Harvey *Scribe: Daniel Busto*

1 Corner Points

We consider three different possible definitions for our intuitive concept of a "corner point". Note that in all of these definitions $P = \{ x : a_i^\intercal x \leq b_i \ \forall i \} \subset \mathbb{R}^n$.

Definition 1.1 (Vertex). x is a **vertex** of a polyhedron P if $\exists c \in \mathbb{R}$ such that $c^\intercal x > c^\intercal y$ $\forall y \in P \setminus \{x\}$

Definition 1.2 (Extreme Point). x is an **extreme point** of a polyhedron P if $\forall y, z \in \mathbb{R}^n$ $\alpha \in (0,1)$ such that $x = \alpha y + (1-\alpha)z$, either $y \notin P$ or $z \notin P$

Definition 1.3 (Basic Feasible Solution). Let $\mathcal{I}_x = \{ i : a_i^\intercal x = b_i \}$, and $A_x = \{ a_i : i \in \mathcal{I}_x \}$. Then x is a **basic feasible solution** if rank $A_x = n$.

Claim 1.4. If $\text{rank}(A_x) < n$, $\exists w \in \mathbb{R}^n$ $w \neq 0$ such that $\forall a_i \in A_x$ $a_i^\intercal w = 0$.

Proof. Let M be the matrix whose i^{th} row is a_i. Recall $\dim(M) = \dim(\text{rowspace}(M)) + \dim(\text{nullspace}(M))$, hence since $\dim(\text{rowspace}(M)) = \text{rank}(A_x) < n$, $\exists w \in \text{nullspace}(M)$. This w is orthogonal to all of A_x. □

Lemma 1.5. If P is a polyhedron the above definitions are equivalent.

Proof. First we show that if x is a **vertex** then it must be an **extreme point**.
Assume x is a **vertex**, by definition $\exists c$ such that $c^\intercal x > c^\intercal y$ $\forall y \in P \setminus \{x\}$.
Let $y, z \in \mathbb{R}^n \setminus \{x\}$ be arbitrary vectors and $\alpha \in (0,1)$ an arbitrary scalar such that $x = \alpha y + (1-\alpha)z$.
Assume $y, z \in P$

$$\begin{aligned} c^\intercal x &= \alpha c^\intercal y + (1-\alpha) c^\intercal z \\ &< \alpha c^\intercal x + (1-\alpha) c^\intercal x \\ &= c^\intercal x \end{aligned}$$

Leading us to conclude $c^\intercal x < c^\intercal x$, which is a contradiction, so either $y \notin P$ or $z \notin P$. Thus x is an **extreme point**.
Second we show that if x is an **extreme point** then x is a **basic feasible solution**.
Assume x is an **extreme point**, and assume to the contrary that it is not a **basic feasible solution**. This means $\text{rank}(A_x) < n$ (where A_x is defined as above).
Recall $\exists w \in \mathbb{R}^n$ $w \neq 0$ such that $\forall a_i \in A_x$ $a_i^\intercal w = 0$.
Let $y = x + \varepsilon w$, $z = x - \varepsilon w$.
We show that ε can be chosen such that $y, z \in P$. Start with $y \in P$.
Consider a constraint a_i such that $a_i^\intercal x = b_i$, then $a_i^\intercal w = 0$ from our choice of w and hence $a_i^\intercal y = a_i^\intercal x + a_i^\intercal w = b_i$.

Next consider the other constraints, those for which $a_i^\intercal x < b_i$. If we take $\varepsilon = \frac{b_i - a_i^\intercal y}{2a_i^\intercal w}$, then we get

$$\begin{aligned} b_i - a_i^\intercal y &= b_i - a_i^\intercal x - \varepsilon a_i^\intercal w \\ &= b_i - a_i^\intercal x - \frac{b_i - a_i^\intercal x}{2} \\ &= \frac{b_i - a_i^\intercal x}{2} \\ &> 0 \\ a_i^\intercal y &< b_i \end{aligned}$$

Which shows that y satisfies the constraint as well.
Since y satisfies the constraints that are satisfied with equality at x, and those satisfied by inequality, it satifies all the constraints of P, that is $y \in P$.
The proof that $z \in P$ is similar, and hence $y, z \in P$. We have that $y, z \in P$ such that $x = \frac{y}{2} + \frac{z}{2}$, and thus x is not an **extreme point**, leading to a contradiction. Hence if x is an **extreme point** it must also be a **basic feasible solution**.
The last step is to show that if x is a **basic feasible solution** then x is a **vertex**.
Assume x is a **basic feasible solution**, that is $\text{rank}(A_x) = n$.
Let $c = \sum_{i \in \mathcal{I}_x} a_i$.
Then

$$\begin{aligned} c^\intercal x &= \sum_{i \in \mathcal{I}_x} a_i^\intercal x \\ &= \sum_{i \in \mathcal{I}_x} b_i \end{aligned}$$

We now show that x is the unique optimizer for c, which is equivalent to x is a vertex.
Let $y \in P$. Then $a_i^\intercal y \leq b_i$.
Hence

$$\begin{aligned} c^\intercal y &= \sum_{i \in \mathcal{I}_x} a_i^\intercal y \\ &\leq \sum_{i \in \mathcal{I}_x} b_i \\ &= c^\intercal x \end{aligned}$$

Thus $\forall y \in P \; c^\intercal x \geq c^\intercal y$. (1)
Let $y \in P$ such that $c^\intercal x = c^\intercal y$.
If $\exists a_i \in A_x$ such that $a_i^\intercal y < b_i$ then $c^\intercal y < c^\intercal x$, so there can't be any such i.
Thus $\forall a_i \in A_x \; a_i^\intercal y = b_i$.
Since $rank(A_x) = n$ there is a unique solution, and hence $x = y$.
Ergo $\forall y \in P$ such that $c^\intercal x = c^\intercal y$ then $x = y$. (2)
Combining (1) and (2) we get $\forall y \in P \setminus \{x\} \; c^\intercal x > c^\intercal y$, which is our definition for x is a **vertex**.
Since we have established:
x is a **vertex** $\implies$ x is an **extreme point** $\implies$ x is a **basic feasible solution** $\implies$ x is a **vertex**
We conclude that all these statements are equivalent. $\square$

Corollary 1.6. Any polyhedron has finitely many extreme points.

Proof. Any polyhedron can be described by $m \in \mathbb{Z}$ constraints, thus there are at most $\binom{m}{n}$ ways to choose constraints to be satisfied by the basic feasible solution, and thus finitely many such points. Since every extreme point is a basic feasible solution, there are no more extreme points than there are basic feasible solutions. Thus there are finitely many extreme points. $\square$

Now that we have discussed common ways to define corner points of polyhedra we move on to showing that optimal solutions of linear programs are realized at extreme points. There are some cases where this isn't the case, specifically when the polyhedra we are optimizing over contains a line (see below for formal definition), or is infeasible. For example the linear program $\min\{\, y \,:\, x \in \mathbb{R}, y \geq 0 \,\}$ has infinitely many optimal solutions (of the form $(x, 0)$), but has no extreme points.

Definition 1.7 (Line). A **line** is any subset L of $\mathbb{R}^n$ of the form $L = \{\, v \,:\, \exists \lambda \in \mathbb{R}\ v = u + \lambda w \,\}$, where $u, w \in \mathbb{R}^n,\ w \neq 0$.

Lemma 1.8. Let $P = \left\{\, x \,:\, \forall i\ a_i^\intercal x \leq b_i \,\right\}$, further assume P does not contain any lines.
Let $\mathcal{L} = \max\left\{\, c^\intercal x \,:\, x \in P \,\right\}$ be a linear program, which has an optimal solution.
Then there is an extreme point of P which is an optimal solution of $\mathcal{L}$.

Proof. Suppose x is an optimal point of $\mathcal{L}$ with the maximal number of tight constraints, that is $\forall x'$ an optimal solution of $\mathcal{L}$ $|\mathcal{I}_{x'}| \leq |\mathcal{I}_x|$. Further assume x is not an extreme point.
Then x is not a basic feasible solution, and as we've seen above $\exists w \in \mathbb{R}^n\ w \neq 0$ such that $\forall i \in \mathcal{I}_x\ a_i^\intercal w = 0$.
Define $y(\varepsilon) = x + \varepsilon w$. Now we have two cases to consider, either $c^\intercal w = 0$ or not.
First assume $c^\intercal w = 0$.
Let $\delta = \min_{i \notin \mathcal{I}_x} \frac{b_i - a_i^\intercal x}{a_i^\intercal w}$, $h = \arg\min_{i \notin \mathcal{I}_x} \frac{b_i - a_i^\intercal x}{a_i^\intercal w}$.
Then $\forall i \in \mathcal{I}_x\ a_i^y(\delta) = a_i^\intercal x = b_i$, since $a_i^\intercal w = 0$, and hence $i \in \mathcal{I}_{y(\delta)}$. For $i \notin I_x$ we have

$$
\begin{aligned}
a_i^\intercal y(\delta) &= a_i^\intercal x + \delta a_i^\intercal w \\
&\leq a_i^\intercal x + a_i^\intercal w \frac{b_i - a_i^\intercal x}{a_i^\intercal w} \\
&= b_i
\end{aligned}
$$

Thus $y(\delta) \in P$, and

$$
\begin{aligned}
a_h^\intercal y(\delta) &= a_h^\intercal x + \delta a_h^\intercal w \\
&= a_h^\intercal x + a_h^\intercal w \frac{b_h - a_h^\intercal x}{a_i^h w} \\
&= b_h
\end{aligned}
$$

Ergo $h \in \mathcal{I}_{y(\delta)}$, as such $|\mathcal{I}_{y(\delta)}| > |\mathcal{I}_x|$.
Finally

$$c^\intercal y(\delta) = c^\intercal x + \delta \underbrace{c^\intercal w}_{=0} = c^\intercal x$$

so $y(\delta)$ is an optimal point of $\mathcal{L}$.
This means $y(\delta)$ is an optimal point of $\mathcal{L}$, for which $|\mathcal{I}_{y(\delta)}| > |\mathcal{I}_x|$, which contradicts our assumption,
Next we consider the case where $c^\mathsf{T} w \neq 0$. Note $\forall i \in \mathcal{I}_x a_i^\mathsf{T}(-w) = -a_i^\mathsf{T} w = 0$ and $c^\mathsf{T}(-w) = -c^\mathsf{T}$. Thus without loss of generality we can assume $c^\mathsf{T} w > 0$, since we can reverse the inequality by replacing w with $-w$.
Let δ be defined as above. Then $y(\delta) \in P$), and

$$c^\mathsf{T} y(\delta) = c^\mathsf{T} x + \delta c^\mathsf{T} w > c^\mathsf{T} x$$

which contradicts x being optimal.
Since we reach a contradiction in both cases the optimal point with maximum number of tight constraints must be an extreme point. □

This leads to a very simple algorithm for solving linear programs. First you go through the $\binom{m}{n}$ possible choices of n constraints, and check which ones give you a basic feasible solution. Then you return the optimal solution out of this set of points.

2 Polyhedrons

We start by defining the dimension of sets.

Definition 2.1. An **affine space** A is a set $A = \{\, x + z \,:\, x \in L \,\}$ where L is a linear space and z is any vector.
The dimension of an affine space, is the dimension of the underlying linear space L.

Definition 2.2. The **dimension** of any $S \subset \mathbb{R}^n$, is $\dim S = \min\{\, A \,:\, A \text{ an affine space } S \subseteq A \,\}$. By convention $\dim \emptyset = -1$.

Definition 2.3. A set C is **convex** if $\forall x, y \in C \; \forall \alpha \in [0,1] \; \alpha x + (1-\alpha) y \in C$

Claim 2.4. The following are convex sets:

1. Closed halfspaces, that is sets of the form $\left\{\, x \,:\, a^\mathsf{T} x \leq b \,\right\}$, where $a \in \mathbb{R}^n$ $b \in \mathbb{R}$.
2. The intersection of a family of convex sets.
3. Polyhedrons.

Closed halfspaces are convex. Let H be the closed halfspace $\left\{\, x \,:\, a^\mathsf{T} x \leq b \,\right\}$, for some $a \in \mathbb{R}^n$ $b \in \mathbb{R}$. Take $y, z \in H$, arbitrary.
Then given $\alpha \in [0,1]$

$$\begin{aligned} a^\mathsf{T}(\alpha y + (1-\alpha) z) &= \alpha a^\mathsf{T} y + (1-\alpha) a^\mathsf{T} z \\ &\leq \alpha b + (1-\alpha) b \\ &= b \end{aligned}$$

so $\alpha y + (1-\alpha) z \in H$, an hence it is convex. □

Intersections of convex families are convex. Let $(C_i)_{i\in I}$ be a family of convex sets, and let $y, z \in \cap_{i\in I} C_i$.
Given $\alpha \in [0,1]$ then $\forall i \in I$ $\alpha y + (1-\alpha)z \in C_i$, since $x, y \in C_i$ which is convex, hence $\alpha y + (1-\alpha)z \in \cap_{i\in I} C_i$. So $\cap_{i\in I} C_i$ is a convex set. □

Polyhedrons are convex. Polyhedrons are the intersections of closed halfspaces. Since we have shown closed halfspaces are convex, and the intersections of convex sets are convex, polyhedrons are convex. □

Definition 2.5. Let $C \subseteq \mathbb{R}^n$ a convex set. An inquality $a^\mathsf{T} x \leq b$ is **valid for C** if $\forall x \in C$ $a^\mathsf{T} x \leq b$.

Definition 2.6. Let $P \subseteq \mathbb{R}^n$ be a polyhedron. A **face** of P is a set $F = P \cap \{ x : a^\mathsf{T} x = b \}$, such that $a^\mathsf{T} x \leq b$ is valid for P.

Claim 2.7. Every face of a polyhedron is a polyhedron.

Proof. Note that $\{ x : a^\mathsf{T} x = b \} = \{ x : a^\mathsf{T} x \leq b \} \cap \{ x : -a^\mathsf{T} x \leq -b \}$. Hence any face of a polyhedron is the intersection of closed halfspaces, and thus a polyhedron. □

Claim 2.8. For any given polyhedron P, P and $\emptyset$ are faces of P.

Proof. Take $a = 0,\ b = 0$ in the above definition. $0^\mathsf{T} x \leq 0$ is true for all x, so it is valid for P. $P \cap \{ x : 0^\mathsf{T} x = 0 \} = P \cap \mathbb{R}^n = P$. Thus P is a face of P. Similarly take $a = 0,\ b = 1$. $0^\mathsf{T} x \leq 1$ is true for all x, and hence valid for P. $P \cap \{ x : 0^\mathsf{T} x = 1 \} = P \cap \emptyset = \emptyset$. Thus $\emptyset$ is a face of P. □

Definition 2.9. A face of a polyhedron with dimension k is called a **k-face**.
An (n-1)-face is a **facet**, a 1-face is an **edge**, a 0-face is a **vertex**.

We now show it makes sense to call 0-faces vertices, by showing it is equivalent to are above definition of a vertex.

Proof. If $F = v$ is a face of some polyhedron, then $\exists a \in \mathbb{R}^n$ $\exists b \in \mathbb{R}$ such that $\forall x \in P$ $a^\mathsf{T} x \leq b$ and $a^\mathsf{T} v = b$. Note that $P \cap \{ x : a^\mathsf{T} x = b \} = v$, so $\forall x \in P \setminus v$ $a^\mathsf{T} x < b$. Hence v is a vertex of P, according to our first definition of a vertex. □

3 Simplex Method

The simplex method can be very simply described as picking an arbitrary starting vertex from the feasible polyhedron, and moving along edges of the polyhedron, until there are no such moves that result in an improvement in the objective value.
This algorithm works very well in practice, and is often preferred to more theoretically complex algorithms which have asymptotically better running times. The exact complexity of the algorithm is unknown because it is hard to find bounds on the shortest to a given vertex on arbitrary polyhedrons.