UBC CPSC 536N: Sparse Approximations Winter 2013

Lecture 11 — February 6, 2013

Prof. Nick Harvey *Scribe: Alexandre Fréchette*

1 Pipage Rounding

Let $P \subseteq \mathbb{R}^m$ be an arbitrary polytope, and let $x \in P$. Moreover, let $f : \mathbb{R}^m \to \mathbb{R}$ be a *linear* function. In Lecture 3 we saw[1] that there is an algorithm that, given $x \in P$, produces an extreme point x^* of P such that $f(x^*) \leq f(x)$.

We want a similar result for a more general class of functions f. To do so, we turn to a similar problem but on the spanning tree polytope.

Definition 1.1. A function $f : \mathbb{R}^m \to \mathbb{R}$ is *partially concave* on a polytope P if, for all $x \in P$ and all $i, j \in [m]$, the univariate function $z \mapsto f(x + z(e_i - e_j))$ is concave.

It turns out that, if P is the spanning tree polytope and f is a partially concave function on P, then there exists a randomized algorithm that, given $x \in P$, produces a (random) extreme point $X^* \in P$ with $\mathbb{E}\{f(X^*)\} \leq f(x)$.

This is our main result for the lecture, and comes in the form of a rounding procedure. We describe the algorithm step-by-step, and provide claims needed to understand it. A lot of notation and concepts are taken from our lecture on the spanning tree polytope.

We start from $x \in P$. If it is already integral, we are done. So assume it is fractional, that is it has a fractional coordinate x_i.

Let $\emptyset = C_0 \subset C_1 \subset C_2 \subset \ldots \subset C_{k-1} \subset C_k = E$ be a maximal chain in $\mathcal{T}_x$. Let C_s be the smallest set in the chain containing i. We know that $x(C_s \setminus C_{s-1}) = x(C_s) - x(C_{s-1}) = r(C_s) - r(C_{s-1})$, which is an integer since $r(C_s)$ and $r(C_{s-1})$ are both integral. Hence, there exists some $j \neq i$ such that $j \in C_s \setminus C_{s-1}$ and x_j is also fractional.

Let $d = e_i - e_j$. Consider moving x in the direction d, that is, along the line $\{x + zd : z \in \mathbb{R}\}$.

Claim 1.2.

$$\mathcal{T}_x \subseteq \mathcal{T}_{x+zd}.$$

In other words, any set tight at x is also tight at $x + zd$.

Proof. Since the C_l's form a chain, every C_l either contains both i and j or neither. Hence, $d^T \chi_{C_l} = 0$ for all $l \in [k]$. Moreover, every $S \in \mathcal{T}_x$ can be written as $\chi_S = \sum_{l \in [k]} \alpha_l \chi_{C_l}$ since $\chi_S \in \mathsf{span}(\{\chi_{C_l} : l \in [k]\})$ for all $S \in \mathcal{T}_x$ (see Lemma 2.7 in Lecture 10). Thus,

$$d^T \chi_S = d^T \sum_{l \in [k]} \alpha_l \chi_{C_l} = \sum_{l \in [k]} \alpha_l d^T \chi_{C_l} = 0.$$

Consider now the point $x_z = x + zd$. Then for any $S \in \mathcal{T}_x$,

$$x_z(S) = x_z^T \chi_S = (x + zd)^T \chi_S = x^T \chi_S + zd^T \chi_S = x^T \chi_S = r(S),$$

[1] This is much simpler than solving a linear program from scratch. In particular, the ellipsoid method is not required. The proof of Lemma 1.8 in Lecture 3 describes the algorithm.

so S is tight for x_z. $\square$

Let now $z^+ = \max\{z \in \mathbb{R} : x + zd \in P\}$ and $z^- = \min\{z \in \mathbb{R} : x + zd \in P\}$.

Claim 1.3. $z^+ > 0$ and $z^- < 0$.

Proof. TODO. $\square$

Let now $p^+ = \frac{-z^-}{z^+ - z^-}$ and $p^- = \frac{z^+}{z^+ - z^-}$. Let Z be a random variable where

$$Z = \begin{cases} z^+ & \text{with probability } p^+ \\ z^- & \text{with probability } p^- \end{cases}.$$

Note that $\mathbb{E}\{Z\} = z^+ \frac{-z^-}{z^+ - z^-} + z^- \frac{z^+}{z^+ - z^-} = 0$.

Let $X_1 = x$ and $X_2 = x_1 + Zd$, and consider the univariate function $z \mapsto f(x + zd)$. We know this function is concave. Recall Jensen's inequality,

Theorem 1.4 (Jensen's Inequality). Given a random variable $X \in \mathbb{R}^m$ and a concave function $g : \mathbb{R}^m \to \mathbb{R}$,

$$\mathbb{E}\{g(X)\} \leq g(\mathbb{E}\{X\}).$$

In our case, since $X_1 = x$ is not random, this means

$$\mathbb{E}\{f(X_2)\} = \mathbb{E}\{f(X_1 + Zd)\} \leq f(\mathbb{E}\{X_1 + Zd\}) = f(X_1 + \mathbb{E}\{Z\}d) = f(x).$$

We then repeat this procedure as long as X_t is fractional, getting a sequence of points $\{X_t\}_{t=1}^k$ that stops when X_k is an integral extreme point. We naturally return $x^* = X_k$.

See Algorithm 1.1 for the full pseudo-code of our proposed method. There are a few caveats and observations we want to make.

First, does the algorithm even terminate? If so, how many times should we repeat the procedure? We note that $\mathcal{T}_{X_i} \subsetneq \mathcal{T}_{X_{i+1}}$ — there is a tight set A in $\mathcal{T}_{X_{i+1}}$ that is not in $\mathcal{T}_{X_i}$. This is clear, because otherwise we could have moved X_i further beyond z^+ or z^-. Thus, the algorithm will terminate, as we will eventually have $\mathcal{T}_{X_k} = \wp(E)$. In particular, the number of iterations is $k \leq |\wp(E)| = 2^m$. In fact, the number of iterations is much less. Note that $\chi_A \notin \mathsf{span}(\{\chi_S : S \in \mathcal{T}_{X_i}\})$ but $\chi_A \notin \mathsf{span}(\{\chi_S : S \in \mathcal{T}_{X_{i+1}}\})$ — that is, the *dimension* of $\{\chi_S : S \in \mathcal{T}_{X_i}\}$ increasing with i. Hence, the number of iterations is $k \leq m$ as the dimension cannot exceed m.

Second, we saw that $\mathbb{E}\{f(X_2)\} \leq f(X_1) = f(x)$. Intuitively, an inductive argument should show that $\mathbb{E}\{f(X_k)\} \leq f(x)$. However, formalizing this induction requires some care with conditional expectations, together with Jensen's inequality for conditional expectation [Grimmett & Stirzaker, Ex. 7.9.4vi] [Klenke, Theorem 8.19]. Alternatively, one can observe that $(X_1, X_2, \ldots, X_k)$ is a martingale, and so $(f(X_1), f(X_2), \ldots, f(X_k))$ is a super-martingale by the partially concave property of f and known facts about martingales [Grimmett & Stirzaker, Ex. 12.1.6] [Klenke, Theorem 9.35].

Our last two concerns are of computational nature. In the next lecture, we will address the issues of finding an inclusion maximal chain in $\mathcal{T}_{X_i}$, and solving to find z^+ and z^-.

Algorithm 1.1 The Pipage Rounding algorithm. It is important to note that the algorithm does not need to know the function f!

Require: A graph G and a point $x \in P$, where P is the spanning tree polyhedron of G.

Ensure: A (random) extreme point X^*. If f is any partially concave function $f : \mathbb{R}^m \to \mathbb{R}$, we have $\mathbb{E}\{f(X^*)\} \leq f(x)$.

$i \leftarrow 1$
$x_i \leftarrow x$
while x_i is not integral, is not an extreme point **do**
  Let $0 = C_0 \subset C_1 \subset \cdots \subset C_k = E$ be a maximal chain in $\mathcal{T}_{x_i}$.
  Let a be any fractional coordinate of x.
  Let C_s be the smallest set in the chain that contains a.
  Let b be any other fractional coordinate in $C_s \setminus C_{s-1}$.
  $z^+ \leftarrow \max\{z \in \mathbb{R} : x_i + z(e_a - e_b) \in P\}$
  $z^- \leftarrow \min\{z \in \mathbb{R} : x_i + z(e_a - e_b) \in P\}$
  $p^+ \leftarrow \frac{-z^-}{z^+ - z^-}$
  $p^- \leftarrow \frac{z^+}{z^+ - z^-}$
  Let $Z = \begin{cases} z^+ & \text{with probability } p^+ \\ z^- & \text{with probability } p^- \end{cases}$
  $x_{i+1} \leftarrow x_i + Z(e_a - e_b)$
  $i \leftarrow i + 1$
end while
return x_i