Rounding Techniques in Approximation Algorithms

Lecture 2: Independent Randomized Rounding for Congestion

Lecturer: Nathan Klein

1 Recap of Relax and Round: Polyhedral View

Remember the relax and round framework we studied last time.

Relax and Round

As input we get a discrete optimization problem O. Now, we:

1. Model O as an Integer Linear Program (ILP): linear constraints with the requirement that the solution x is integer.
2. **Relax** the condition to $x \in \mathbb{R}^n$, giving us a Linear Program (LP). Solve the resulting LP using the ellipsoid method to obtain a solution x.
3. **Round** the point x to a solution to O.

It is convenient to think about the ILP and the LP as producing solutions to optimization problems over convex polyhedra.

Definition 1.1 (Convex Polyhedron/Polytope)**.** *A convex polyhedron is the intersection of finitely many half-spaces of the form $a^T x \geq b$ for $a \in \mathbb{R}^n, b \in \mathbb{R}$. A convex polytope is a bounded convex polyhedron. See Fig. 1.*

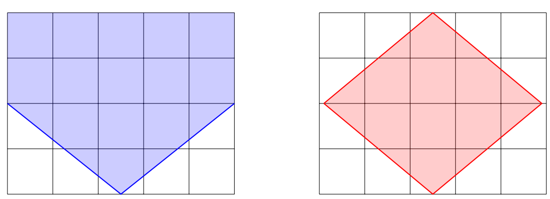

Figure 1: On the left is a convex polyhedron defined by two half-planes, and on the right is a convex polytope defined by four half-planes.

Thus, the set of linear constraints for any ILP form a polyhedron (or a polytope), as do the set of all constraints for an LP. We can therefore think of an ILP/LP pair as follows. First, define a polyhedron P with a separation oracle. Then, the ILP (left) and LP (right) are:

$$\begin{array}{rl} \min & c(x) \\ \text{s.t.} & x \in P \\ & x \in \mathbb{Z}^n \end{array} \qquad\qquad \begin{array}{rl} \min & c(x) \\ \text{s.t.} & x \in P \\ & x \in \mathbb{R}^n \end{array}$$

This highlights the importance of the integrality gap: it gives us the largest possible difference between integer points and real points in a polyhedron. So really, integrality gap is defined with respect to a polyhedron.

Here is an example of the relax-and-round framework geometrically. Suppose the marked points in $\mathbb{Z}^2$ are the feasible solutions to our optimization problem O. Then the blue polytope P is a valid ILP for the problem since $P \cap \mathbb{Z}^2$ is the set of feasible solutions for O.

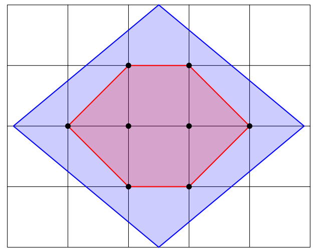

Figure 2: In purple is a polytope P with a separation oracle. In red is the convex hull of the integer coordinates in that polytope. The ILP returns an optimal vertex of the red set. The relaxed LP returns an optimal vertex of the purple set. The integrality gap measures how different these two vertices can be in terms of cost.

If we could find a separation oracle for the red convex set for an NP-Hard problem, we would prove P=NP. So, such a thing likely does not exist, and we have to settle for approximations of the red polytope. Let's redefine the relax-and-round framework with our new language:

Relax and Round: Polyhedral Version

As input we get an instance of a discrete optimization problem O. Now, we:

1. Find a polyhedron P with a polynomial time separation oracle such that $P \cap \mathbb{Z}^n$ is the set of feasible solutions to O.
2. **Relax** our problem: use the ellipsoid method to find an optimal vertex $x \in \mathbb{R}^n$ in P.
3. **Round** the point x to a solution to O.

There are many choices for P, and as we saw in the last lecture, often the most natural choice can be strengthened with additional inequalities. We will explore this further in later lectures.

Notice that in Fig. 2 and in the above definition I say we find an optimal *vertex*. This is because given a polyhedron, if we optimize in any direction, an optimal solution (so long as it is finite) will always be attained at some vertex of that set.

Definition 1.2 (Vertex)**.** *Given a polyhedron P, a point $v \in P$ is a vertex of P if and only if there is no direction $0 \neq d \in \mathbb{R}^n$ such that $v + d \in P$ and $v - d \in P$.*

A nice property of polytopes is the following:

Theorem 1.3 (Carathéodory's Theorem)**.** *Let x lie in the convex hull of a set $S \subseteq \mathbb{R}^n$. Then x can be written as the convex combination of at most $n+1$ points of S.*

It also gives the following fact immediately:

Fact 1.4. *The optimum value of an LP can always be attained at a vertex.*

Proof. Let $a \in P$ be an optimal solution. By Carathéodory, we can write a as a convex combination of vertices. So some vertex has cost at most that of a. ☐

2 Independent Randomized Rounding

Last time we saw how to round the natural LP for Vertex Cover to achieve a 2-approximation using *threshold rounding*, which is optimal under the Unique Games Conjecture. Today, we will look at another very basic rounding technique: *randomized rounding.* Given an LP solution $x \in [0,1]^n$, this technique independently rounds each variable i to 1 with probability x_i and to 0 otherwise for some fixed (usually linear) function $f : [0,1] \to [0,1]$.

2.1 Chernoff Bounds and the Union Bound

Chernoff bounds are immensely important for understanding independent randomized rounding procedures.

Theorem 2.1 (Chernoff Lower Tail)**.** *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables. Then if $X = \sum_{i=1}^n X_i$, $L \leq \mathbb{E}[X]$, and $0 \leq \delta \leq 1$, we have:*

$$\mathbb{P}[X \leq (1-\delta)L] \leq \exp(-L\delta^2/2)$$

Theorem 2.2 (Chernoff Upper Tails)**.** *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables. Let $X = \sum_{i=1}^n X_i$, $\mathbb{E}[X] \leq U$. For all $\delta > 0$, we have*

$$\mathbb{P}[X \geq (1+\delta)U] \leq \exp(-U\delta^2/(2+\delta)).$$

Notice that if $\delta \leq 1$ then we can use $\exp(-U\delta^2/3)$. As δ grows, the following bound is superior and holds for $\delta \geq 2$:

$$\mathbb{P}[X \geq (1+\delta)U] \leq \exp(-U\delta \ln \delta/4)$$

Here's the version these bounds come from, which we often avoid because it's a bit annoying to use, but can be slightly stronger:

Theorem 2.3 (Mother-of-all Chernoff Bounds)**.** *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables. Then if $X = \sum_{i=1}^n X_i$, $\mu = \mathbb{E}[X]$, $L \leq \mu \leq U$, and $\delta > 0$, we have:*

$$\mathbb{P}[X \geq (1+\delta)U] < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^U.$$

and for $0 < \delta < 1$,

$$\mathbb{P}[X \leq (1-\delta)U] < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L$$

Finally I will give one more bound that is useful for variables that are not binary.

Theorem 2.4 (More General Chernoff)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $|X_i| \leq a_i$ for each i. Let $X = \sum_{i=1}^n X_i$. Then if $\delta \geq 0$, we have*

$$\mathbb{P}[|X| \geq \lambda \|a\|_2] \leq 2\exp(-\lambda^2/4)$$

Note that **the lower and upper tails are qualitatively different** in the following way. Using the lower tail for $\delta = 1 - \epsilon$,

$$\mathbb{P}\left[X \leq \epsilon \mathbb{E}\left[X\right]\right] \leq e^{-\mathbb{E}[X](1-\epsilon)^2/2}$$

But using the upper tail for $\delta = \frac{1}{\epsilon}$,

$$\mathbb{P}\left[X \geq \frac{1}{\epsilon}\mathbb{E}\left[X\right]\right] \leq e^{-\mathbb{E}[X]\frac{1}{4\epsilon}\ln(\frac{1}{\epsilon})}$$

Thus, the probability X is less than $\frac{1}{100}$ times your expectation is approximately upper bounded by $e^{-\mathbb{E}[X]}$. The probability X is at least 100 times its expectation is approximately upper bounded by $e^{-100\mathbb{E}[X]}$. As $\epsilon \to 0$, these two quantities become very different. You will see this come up in the second homework.

Chernoff bounds are often coupled with the union bound.

Theorem 2.5 (Union Bound)**.** *Let $B_1, \ldots, B_n$ be a set of bad events over a probability space. Then,*

$$\mathbb{P}\left[\overline{B_1} \wedge \cdots \wedge \overline{B_n}\right] \geq 1 - \sum_{i=1}^{n} \mathbb{P}\left[B_i\right]$$

This is simply because the probability any one of the bad events happens is at most $\sum_{i=1}^{n} \mathbb{P}\left[B_i\right]$, corresponding to the possibility that they all occur at different times.

2.2 A Simple Example of Chernoff/Union Bound: Balls in Bins

Suppose we throw n balls into n bins, each ball going into a bin uniformly at random. Then, we can bound the probability any bin gets more than $12 \ln n$ balls.

The expected number of balls in any one bin is 1. Lazily applying the first Chernoff upper tail for $\delta = 1$, $U = \ln n$, we get:

$$\mathbb{P}\left[\text{Bin } i \text{ has at least } 12 \ln n \text{ balls}\right] \leq \exp(-6 \ln n/3) = \frac{1}{n^2}$$

This creates n bad events of probability at most $\frac{1}{n^2}$ each. So the probability we avoid all of them, by the union bound, is at least $1 - 1/n$.

This can be improved to $O(\log n / \log \log n)$ with a couple more lines of computation.

2.3 Congestion Minimization for Multi-Commodity Flows

To explain randomized rounding we will work on the minimum-congestion multicommodity flow problem. Here we are given a directed graph $G = (V, A)$ and k pairs of vertices $s_i, t_i \in V$ for $1 \leq i \leq k$. The goal is to produce a path from s_i to t_i for each i so that the number of paths using any single edge is minimized.

Let's relax and round. Step 1 is to model the problem as an ILP. The natural one is as follows, where $\mathcal{P}_i$ is the set of simple paths from s_i to t_i in G.

$$\begin{aligned}
\min \quad & \lambda \\
\text{s.t.} \quad & \sum_{P \in \mathcal{P}_i} x_P = 1 && \forall 1 \le i \le k \\
& \sum_{P : e \in P} x_P \le \lambda && \forall e \in E \\
& x_P \in \{0,1\} && \forall P \in \mathcal{P}_i, 1 \le i \le k
\end{aligned}$$

This is equivalent to the original problem. Now let's do Step 2: relax this to an LP, by replacing the requirement that $x_P \in \{0,1\}$ with $0 \le x_P \le 1$. At this point, we need to solve the LP in polynomial time to continue with the framework. But we realize an issue: there are exponentially many variables. It turns out this is easy to circumvent. An equivalent problem is to just find a feasible flow for each i. We can easily encode this as an LP.

$$\begin{aligned}
\min \quad & \lambda \\
\text{s.t.} \quad & \sum_{a \in \delta^+(v)} f_a^i = \sum_{a \in \delta^-(v)} f_a^i && \forall 1 \le i \le k, v \in V, v \ne s_i, t_i \\
& \sum_{a \in \delta^+(s_i)} f_a^i = 1 && \forall 1 \le i \le k \\
& \sum_{i=1}^{k} f_a^i \le \lambda && \forall a \in A \\
& 0 \le f_a^i \le 1 && \forall 1 \le i \le k, a \in A
\end{aligned}$$

Now, we can use this to solve the previous LP (and obtain the same objective value) with a polynomial number of non-zero values x_P by doing a flow decomposition.

Lemma 2.6. *Given a solution f to the flow LP above of value λ, we can construct a solution x in polynomial time to the path LP with polynomially many non-zero x_P.*

Proof. For each flow f^i, find a path P from s_i to t_i using edges with $f_a^i > 0$. Set $x_P = \min_{a \in P} f_a^i$ in $\mathcal{P}_i$. Then, set $f_a^i = f_a^i - x_P$ for all $a \in P$. Repeat until no paths remain using edges with positive flow value. By flow conservation, $\sum_{P \in \mathcal{P}_i} x_P = 1$ at the end of this process. □

We will now work with our solution x that has polynomially many non-zero entries and perform independent randomized rounding. To do so, we will treat each $\mathcal{P}_i$ as a distribution for each i, and we will sample exactly one path such that the probability we take path P for $P \in \mathcal{P}_i$ is exactly x_P. Our output will be the set of paths we pick. We will now argue that with high probability, the congestion is not too large compared to OPT.

Lemma 2.7. *With high probability the congestion is $O(\log n) \cdot \max\{1, \lambda\}$. Thus, with high probability we obtain a $O(\log n)$ approximation on the minimum congestion.*[1]

[1] In the first homework you will sharpen this to $O(\log n / \log \log n)$.

Proof. For each arc a and $1 \leq i \leq k$, we create a random variable X_a^i which is 1 if the path chosen for pair s_i, t_i contains a and 0 otherwise. For each arc let $C_a = \sum_{i=1}^k X_a^i$ denote the congestion on a. Then,

$$\mathbb{E}\left[Y_a\right] = \sum_{i=1}^{k} \sum_{P \in \mathcal{P}_i : a \in P} x_P \leq \lambda$$

Now note that for every arc a, X_a^i is independent of X_a^j for $i \neq j$. So, we can (lazily) apply a Chernoff bound for $U = c \ln n \cdot \lambda'$ (where $\lambda' = \max\{1, \lambda\}$) and $\delta = 1$ for some c we will choose later. Then we have:

$$\mathbb{P}\left[X \geq 2c \ln n \cdot \lambda'\right] \leq e^{-c \ln n \cdot \lambda'/3} = n^{-c\lambda'/3} \leq n^{-c/3}$$

Choose $c \geq 6$. Then, the probability any individual edge has congestion greater than $O(\log n) \cdot \lambda'$ is at most $\frac{1}{n^2}$. Now by the union bound, all edges have congestion at most $O(\log n) \cdot \lambda'$ with probability at least $1 - \frac{1}{n}$. $\square$

Notice this is a situation (as hinted at in the last lecture) in which the integrality gap may be much larger than the claimed approximation ratio! Indeed, even if we have only one s_i, t_i pair, we can potentially spread the congestion out to get a solution of cost $O(\frac{1}{n})$, but clearly we need to have congestion at least 1.

After sharpening this to $O(\log n / \log \log n)$, it turns out this is optimal. There is an integrality gap (with $\lambda \geq 1$) with this value, and it is NP-Hard to do better than this.