Using SDPs to design approximation algorithms.

Symmetric $n \times n$ matrix M with real entries is *positive semidefinite* (psd) if it can be represented as $M = A^T A$ for some $n \times n$ matrix A.

Thinking of the columns of A as n-dim vectors $u_1, u_2, .., u_n$, we see that (i, j) entry of M is $M(i, j) = u_i \cdot u_j$.

General form of Semidefinite Program (SDP): Find $n \times n$ psd matrix X that satisfies $a_s \cdot X \geq b_s$ for s=1, 2, …, m and minimizes $c \cdot X$.

Here a_s and c are n^2-dimensional vectors.

In other words, SDP consists of finding a set of n vectors $u_1, u_2, .., u_n \in \Re^n$ such that the inner-products $u_i \cdot u_j$ satisfy some given linear constraints and you minimize some linear function of the innerproducts.

In HW4 we saw how to solve SDPs in polynomial time using the Ellipsoid algorithm, and a separation oracle for the semidefinite cone.

SDP relaxation for MAX-CUT

Problem: Given $G=(V, E)$ find a cut (S, S^c) that maximizes the number of edges $E(S, S^c)$ in the cut (S, S^c).

We can represent this as the following quadratic integer program:

Find $x_1, x_2, .. x_n \in \{-1, 1\}$ so as to maximize:

$$\sum_{(i,\ k) \in E} \tfrac{1}{4} (x_i - x_k)^2$$

SDP relaxation:

Find $u_1, u_2, \ldots, u_n \in \Re^n$ such that $|u_i|_2 = 1\ \forall\ i$ and to maximize

$$\sum_{(i,\ k) \in E} \tfrac{1}{4} |u_i - u_k|^2$$

Note: (i) This is a relaxation since $\{-1, 1\}$ solutions are still feasible.
(ii) $|u_i|_2 = 1$ is same as $u_i \cdot u_i = 1$ and $|u_i - u_k|^2 = u_i \cdot u_i + u_k \cdot u_k - 2u_i \cdot u_k$

So this fits the SDP paradigm.

Rounding the Max-Cut SDP (Goemans-Williamson '94)

Pick a random unit vector $z \in \Re^n$. Let $S = \{i: u_i \cdot z \geq 0\}$.
Output cut (S, S^c).

Analysis. We are essentially picking a random hyperplane through the origin and partitioning vectors according to which side they lie on. We estimate the probability that an edge (i, k) is in the final cut.

Let Θ_{ik} = angle between u_i, u_k = $\cos^{-1}(u_i \cdot u_k)$

Pr[u_i, u_k are on opp. sides of a random hyperplane] = Θ_{ik}/π

E[# of edges in final cut] = $\sum_{(ik) \in E} \Theta_{i\,k}/\pi$. (*)

SDP value = $\sum_{(i\,k) \in E} \frac{1}{4} |v_i - v_k|^2 = \sum_{(i\,k) \in E} \frac{1}{2} (1 - \cos(\Theta_{i\,k}))$ (**)

Fact: For all $\Theta \in [0,\pi]$, $\Theta/\pi \geq 0.878 \times \frac{1}{2} (1 - \cos(\Theta))$

So (*) is at least 0.878 × (**). Thus this gives a 0.878-approximation.

Integrality gap

The Goemans-Williamson analysis also shows that

MAX-CUT $\geq$ 0.878 $\times$ (Opt. value of SDP relaxation)

We say that the relaxation has *integrality gap* at most 0.878.

(In fact this estimate is tight for some graphs.)

SDP relaxation for C-BALANCED SEPARATOR

Problem: Given $G=(V, E)$ find a cut (S, S^c) whose each side contains at least Cn nodes and that minimizes $|E(S, S^c)|$.

We can represent this as the following quadratic integer program:

Find $x_1, x_2, .., x_n \in \{-1, 1\}$ satisfying $\sum_{i<k} \frac{1}{4} (x_i - x_k)^2$
so as to minimize:

$$\sum_{(i, k) \in E} \frac{1}{4} (x_i - x_k)^2$$

SDP Relaxation: Try 1

Find unit vectors $u_1, u_2, .., u_n \in \Re^n$ satisfying $\sum_{i < k} \frac{1}{4} |u_i - u_k|^2$ and minimizes $\sum_{(i, k) \in E} \frac{1}{4} |v_i - v_k|^2$.

The Goemans-Williamson analysis does not extend to minimization problems Such as this one. The inequality goes the wrong way! We need an *upperbound* on the cost of the output solution, not a *lowerbound*.

In fact, this relaxation has integrality gap $\Omega(n)$.

SDP Relaxation for C-BALANCED SEPARATOR: Try 2

General idea behind stronger relaxations: throw in constraints that are satisfied by the $\{-1, 1\}$ solution.

Obs.: If $x_1, x_2, x_3 \in \{-1, 1\}$ then $(x_1 - x_2)^2 + (x_2 - x_3)^2 \geq (x_1 - x_3)^2$ ("Triangle inequality")

Find unit vectors $u_1, u_2, .., u_n \in \Re^n$ satisfying (a) $\sum_{i<k} \frac{1}{4} |u_i - u_k|^2$,
(b) $\forall$ i, k, l $|u_i - u_k|^2 + |u_k - u_l|^2 \geq |u_i - u_l|^2$ and
Minimizing $\sum_{(i, k) \in E} \frac{1}{4} |u_i - u_k|^2$.

Let OPT_{SDP} = optimum value of this SDP.

Thm (Arora, Rao, Vazirani'04) There is a randomized rounding algorithm that uses these vectors to produce a C'-balanced cut whose capacity is at most $O(\sqrt{\log n}) \times OPT_{SDP}$. Here $C' = C/5$.

Main Lemma: For some $\Delta = \Omega(1/\sqrt{\log n})$ there is randomized poly-time algorithm that, given any vectors $u_1, u_2, .., u_n$ satisfying the SDP constraints, finds sets S, T of C'n vectors each such that for each $u_i \in S, u_k \in T$:

$|u_i - u_j|^2 \geq \Delta$. (We say such sets are "Δ-separated.")

(Aside: in this Lemma, Δ is best possible up to a constant factor.)

Claim: These sets allow us to find a C'-balanced cut of capacity $\leq SDP_{OPT}/\Delta$

Proof: Define $d(u_i, u_k)$ as $|u_i - u_k|^2$. Note: this satisfies triangle inequality. Define $d(u_i, S) = \min_{u\{k\} \in S} \{d(u_i, u_k)\}$.

For each $x \in [0,\Delta)$ consider the cut (S_x, S_x^c) where $S_x = \{i: d(u_i, S) \leq x\}$, and take the cut of minimum capacity among all such cuts. (Note: we only Need to compare at most n distinct cuts.) Note that each such cut is C'-balanced since S and T are on opposite sides. Let K be the capacity of the output cut.

We show $\sum_{(i, k) \in E} d(u_i, u_k) \geq K \Delta$, which will prove the Claim.

Why $\sum_{(i,\ k) \in E} d(u_i, u_k) \geq K\,\Delta$.

Consider $\int_0^\Delta |E(S_x, S_x^c)|\ dx$.

The function being integrated has value at least K in the entire interval, so the value is at least $K\,\Delta$.

On the other hand, an edge (u_i, u_k) contributes to this integral only when $x \in [d(u_i, S), d(u_k, S))$.

Triangle inequality implies that $d(u_k, S) - d(u_i, S) \leq d(u_i, u_k)$.

We conclude that the value of the integral is at most $\sum_{(i,\ k) \in E} d(u_i, u_k)$.

QED.

Proof of [ARV] Theorem on Δ-separated sets (baby version)

We prove for $\Delta = 1/\log n$ and $c' = c/100$ or so ($\Delta = 1/\sqrt{\log n}$ is harder!)

Poincare's Lemma: *Let d be large, v :arbitrary unit vector $\in \Re^d$ and z: random unit vector.*

Then the real-valued random variable $z \cdot v \sqrt{d}$ is distributed essentially like the gaussian with mean 0 and standard deviation 1.

(Intuition:Mean =0 is clear since $z \cdot v$ is distributed symmetrically about 0; the distributions z and –z are identical. The gaussian behavior is deduced by a simple volume computation.)

Claim: Foll. algorithm produces Δ-sep. sets whp. Pick a random unit vector z and compute $u_1 \cdot z, u_2 \cdot z, .., u_n \cdot z$. Let ε be a suitable constant that depends on c. Let $S = \{i: u_i \cdot z < -\varepsilon/\sqrt{d}\}$, $T = \{j: u_j \cdot z > \varepsilon/\sqrt{d}\}$.

Proof sketch:

<u>Claim (i)</u> Whp, $\forall$ i, k $|\langle u_i - u_k, z\rangle| \le O(\sqrt{\log n}) \times |u_i - u_k|_2/\sqrt{d}$.

Reason: The probability that a Gaussian random variable takes a value that is s standard deviations away from the mean is roughly exp $(-s^2/2)$. Here we are interested in $s = O(\sqrt{\log n})$, which happens with probability $\exp(-\log n) = n^{-c}$. The number of pairs (i, k) is at most n^2, so the union bound implies the claim.

Now note that if $u_i \in S$, $u_k \in T$ then $(u_k - u_i) \cdot z \ge 2\varepsilon/\sqrt{d}$. If the event in Claim (i) happens, then we have $|u_i - u_k|_2 \ge \Omega(1/\log n)$, and so S, T are Δ-separated for $\Delta = \Omega(1/\log n)$.

<u>Claim (ii)</u> Probability $|S|, |T| \ge c' n$ is $\Omega(1)$. (Constt c' to be determined.)

Let $S_z = \{i: \langle u_i, z\rangle \ge 0\}$. Note: this is exactly the Goemans-Williamson cut. Hence $E[\sum_{i<k} 1] \ge 0.878 \times \sum_{i<k} \frac{1}{4} |u_i - u_k|^2 \ge 0.878 \times c(1-c)n^2$.

Averaging implies that with probability 0.5, this quantity is $\Omega(n^2)$, in which case both S_z, and its complement have $\Omega(n)$ nodes.

Finally, Gaussian behavior of projections $\Rightarrow$ With prob 0.99, $|S| \ge |S_z| - 100\varepsilon n$. Making ε small enough, we conclude that with probability > 0.49, $|S| = \Omega(n)$.