Cornell University, Fall 2018	**CS 6820: Algorithms**
Lecture notes	**Prepared by Jonathan Shi**
MAX-CUT and semidefinite programming	**October 10, 2018**

These notes on semidefinite programming will differ a bit from the usual presentation. Recent research has tackled a family of semidefinite programming relaxations known as the Sum-of-Squares hierarchy from the perspective of statistics, and made much progress on statistical inference problems in doing so. These analyses are based loosely on the *method of moments* in statistics, where empirical estimates of the *moments* ($\mathbb{E}\, p(x)$ for monomials $p(x)$) of an unknown distribution are used to estimate the parameters of that distribution.

This motivates us to frame semidefinite programming itself in the language of statistics. First we introduce the MAX-CUT problem, and discussing how linear programming methods are insufficient to solve (or even approximate) this problem. After that, we'll define semidefinite programs and analyze their application to MAX-CUT via the Goemans-Williamson algorithm.

Contents

1 The MAX-CUT problem

Definition 1.1. *A* cut *of an undirected unweighted graph $G = (V, E)$, is a set $A \subseteq V$, whose* capacity *is*

$$c(A) := \left| \left\{ \{u, v\} \in E : \ u \in A \ \oplus \ v \in A \right\} \right|,$$

where $\oplus$ is the exclusive OR.

Problem 1.2. MAX-CUT*: Given an undirected graph $G = (V, E)$, find a cut $A \subseteq V$ that satisfies*

$$c(A) = \max_{U \subseteq V} \ c(U) \, .$$

We don't know how to solve this problem in polynomial-time, and in fact it is NP-complete. However, progress has been made on finding *approximately* optimal solutions to this problem.

Problem 1.3. α-approximate MAX-CUT*: Given an undirected graph $G = (V, E)$, find a cut $A \subseteq V$ that satisfies*

$$c(A) \geq \alpha \max_{U \subseteq V} \ c(U) \, .$$

We'll observe that a 0.5-approximation is readily obtained by simply taking a random cut of the graph. So our attention will be focused on improving this ratio.

Proposition 1.1. *A random cut $A \subseteq V$, where each $v \in V$ is in A with 50% probability and not in A with 50% probability, attains a* 0.5*-approximate* MAX-CUT*.*

Proof. Use the linearity of expectation to find the expected value of $c(A)$ as the sum of the probabilities that each edge in E is cut. Then the 0.5-approximation follows by Markov's inequality, taking repeated trials if necessary. Letting $\mathbb{1}[p]$ denote the

function that is 1 when p is a true proposition and 0 when p is false:

$$\begin{aligned}
\mathop{\mathbb{E}}_{A} c(A) &= \mathop{\mathbb{E}}_{A} \sum_{\{u,v\}\in E} \mathbb{1}\left[u \in A \ \oplus \ v \in A\right] \\
&= \sum_{\{u,v\}\in E} \mathop{\mathbb{E}}_{A} \ \mathbb{1}\left[u \in A \ \oplus \ v \in A\right] \\
&= \sum_{\{u,v\}\in E} \Pr_{A}\left[u \in A \ \oplus \ v \in A\right] \\
&= \sum_{\{u,v\}\in E} \tfrac{1}{2} \\
&= \tfrac{1}{2}|E| \\
&\geq \tfrac{1}{2} \max_{U\subseteq V} \ c(U)\,.
\end{aligned}$$

□

2 Attempt by linear programming (LP)

We may write MAX-CUT as the following mathematical program:

$$\begin{aligned}
\max_{x} \quad & \sum_{\{u,v\}\in E} \tfrac{1}{2}(1 - x_u x_v) \\
\text{s.t.} \quad & x_v \in \{-1, 1\}\,, \qquad\qquad \forall v \in V\,.
\end{aligned}$$

We interpret x_v as indicating which side of the cut v is in. Then $\frac{1}{2}(1 - x_u x_v)$ is 0 if u and v are on the same side of the cut, and 1 if different.

So try a linear relaxation... first we create $z_{\{u,v\}}$ as a stand-in variable for $x_u x_v$. Then for $x_u, x_v, z_{\{u,v\}} \in \{-1, 1\}$, the condition $z_{\{u,v\}} \geq x_u x_v$ is captured by the linear inequalities $z_{\{u,v\}} \geq -x_u - x_v - 1$ and $z_{\{u,v\}} \geq x_u + x_v - 1$.

$$\begin{aligned}
\max_{x,z} \quad & \sum_{e\in E} \tfrac{1}{2}(1 - z_e) \\
\text{s.t.} \quad & z_{\{u,v\}} \geq -x_u - x_v - 1\,, \quad \forall \{u, v\} \in E\,, \\
& z_{\{u,v\}} \geq \ \ x_u + x_v - 1\,, \quad \forall \{u, v\} \in E\,, \\
& x_v \in \{-1, 1\}\,, \qquad\qquad \forall v \in V\,.
\end{aligned}$$

Now we relax the integer constraints to linear ones:

$$\begin{aligned}
\max_{x,z} \quad & \sum_{e \in E} \tfrac{1}{2}(1 - z_e) \\
\text{s.t.} \quad & z_{\{u,v\}} \geq -x_u - x_v - 1\,, \quad \forall \{u,v\} \in E\,, \\
& z_{\{u,v\}} \geq \quad x_u + x_v - 1\,, \quad \forall \{u,v\} \in E\,, \\
& -1 \leq x_v \leq 1\,, \qquad\qquad \forall v \in V\,.
\end{aligned}$$

But notice ($\forall v \in V\,.\, x_v = 0$) and ($\forall e \in E\,.\, z_e = -1$) is a solution that "cuts" every edge!

This difficulty will occur in every LP formulation, due to *symmetry*: whenever A is a cut achieving good value, then the complement of A is also a cut achieving good value. Even if you tried to "break the symmetry" by adding local constraints, say by fixing $x_{v_0} = 1$ for some $v_0 \in V$, you could still flip the sign of every other coordinate of x and still take only a small $\deg(v_0)$-sized loss in the value of the solution. Therefore, any relaxation where the average of two good solutions is also a good solution *cannot work* for this problem!

We need the quadratic term $x_u x_v$ that we tried to eliminate by introducing the variable z_e!

2.1 Interpretation as expectations and marginal probabilities

In the homework, we've seen LPs with variables bounded between 0 and 1 that were interpreted as marginal probabilities. Even when LP variables are not bounded between 0 and 1, they may be interpreted as *expected values*: every fractional solution to an LP relaxation of an integer program is a convex combination of integer-valued solutions. In the same way, the expected values of a probability distribution over those integer-valued solutions is a convex combination of those solutions.

This interpretation will guide our generalization for adding quadratic terms to LPs. First we will understand *covariance matrices*—equivalently, the expected values of quadratic terms.

3 Covariance and positive-semidefinite matrices

Definition 3.1. *The* covariance matrix *of a distribution D over $\mathbb{R}^n$ is defined as* $\mathbb{E}_{x\sim D}\, xx^T$.

Observe that knowing the covariance matrix tells you the expected value of every homogeneous degree-2 polynomial. Each such polynomial is a linear combination of degree-2 monomials, and each degree-2 monomial x_ix_j has expectation $\mathbb{E}\, x_ix_j = \mathbb{E}\, \langle e_i, x\rangle \langle e_j, x\rangle = \mathbb{E}\, e_i^T xx^T e_j = e_i^T(\mathbb{E}\, xx^T)e_j$.

We will need tools to analyze covariance matrices. The first of these is the eigendecomposition, and the accompanying spectral theorem, stated here without proof.

Theorem 3.1 (The Spectral Theorem)**.** *Any symmetric (or, more generally, normal) matrix M has an eigendecomposition $M = V\Lambda V^T$, where Λ is a diagonal matrix of* eigenvalues *λ_i, and V is an orthogonal matrix whose columns are their corresponding (unit-length and mutually orthogonal)* eigenvectors *v_i. This implies that $Mv_i = \lambda_i v_i$.*

In particular, a matrix is the covariance matrix of some distribution if and only if it has all-nonnegative eigenvalues, as we will see.

Definition 3.2. *A* positive-semidefinite (PSD) matrix *M is a symmetric matrix whose eigenvalues are all non-negative. We will denote M being positive-semidefinite by writing $M \succeq 0$ (in this notation, referred to as the* Loewner order*, $A \preceq B$ means that $B - A \succeq 0$).*

Proposition 3.2. *For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, the following statements are equivalent:*

1. $M \succeq 0$.
2. $M = AA^T$ *for some matrix* A.
3. $v^TMv \geq 0$ *for all* $v \in \mathbb{R}^n$.

Proof. **(1) $\Longrightarrow$ (2):** By definition $M \succeq 0$ implies $M = V\Lambda V^T$ where Λ is diagonal with nonnegative entries. Define $\Lambda^{1/2}$ as the diagonal matrix so that $(\Lambda^{1/2})_{i,i} = \sqrt{\Lambda_{i,i}}$. Then, taking $A = V\Lambda^{1/2}$, we find $M = V\Lambda^{1/2}\Lambda^{1/2}V^T = AA^T$.

(2) $\Longrightarrow$ (3): If $M = AA^T$, then $v^T M v = v^T A A^T v = \langle Av, Av \rangle \geq 0$.

(3) $\Longrightarrow$ (1): If $v^T M v \geq 0$ for all $v \in \mathbb{R}^n$, then in particular, $v_i^T M v_i \geq 0$ for each eigenvector v_i of M. With λ_i its eigenvalue then, $Mv_i = \lambda_i v_i$, so that $v_i^T \cdot \lambda_i v_i = \lambda_i \geq 0$. So all eigenvalues are non-negative.

□

Proposition 3.3. *$M \succeq 0$ if and only if M is the covariance matrix of some distribution over n variables. That is, if and only if there is a distribution D over $\mathbb{R}^n$ such that $M = \mathbb{E}_{x \sim D}\, xx^T$.*

Proof. By Proposition 3.2, if $M \succeq 0$ then $M = AA^T$ for some A. If $M = AA^T$, then we may take as our distribution D the uniform distribution over times the columns $a_1, \ldots, a_n$ of A multiplied by $\sqrt{n}$. Then $\mathbb{E}_{x \sim D}\, xx^T = \frac{1}{n} \sum_{i \in [n]} n\, a_i a_i^T = AA^T$.

If M is the covariance matrix of a distribution D, then by linearity of expectation, $v^T M v = \mathbb{E}_{x \sim D}\, v^T x x^T v = \mathbb{E}_{x \sim D} \langle x, v \rangle^2$. The expectation of an always-nonnegative quantity must itself be nonnegative, so $v^T M v \geq 0$ for all vectors v. By Proposition 3.2, this is sufficient to conclude $M \succeq 0$. □

In fact, taking any distribution D with a covariance of the identity matrix and multiplying the samples $x \sim D$ by $M^{1/2}$ (the unique positive-semidefinite matrix satisfying $(M^{1/2})^2 = M$) yields a random variable $y = M^{1/2} x$ whose covariance matrix is M. We can check this: $\mathbb{E}\, yy^T = \mathbb{E}_{x \sim D}\, M^{1/2} x x^T M^{1/2} = M^{1/2} (\mathbb{E}_{x \sim D}\, xx^T) M^{1/2} = M$.

Examples of distributions with covariance Id include

- the uniform distribution over the radius-$\sqrt{n}$ sphere and
- the standard Gaussian $\mathcal{N}(0, \mathrm{Id})$.

The Gaussian distribution $\mathcal{N}(0, M)$ may be defined as $M^{1/2}$ times $\mathcal{N}(0, \mathrm{Id})$. As a side observation, transforming a sphere with $M^{1/2}$ results in an ellipsoid.

4 Semidefinite programming

We'll use $\mathbb{E}_{x \backsim M}$ to denote the expectation (of a degree-2 homogeneous polynomial) over x, when x is drawn from a distribution with covariance M.

The generic form of a semidefinite program (SDP) may be written, for degree-2 homogeneous polynomials $p(x)$ and $q_i(x)$ and $b_i \in \mathbb{R}$,

$$\begin{aligned} \min_{M \succeq 0} \quad & \mathop{\mathbb{E}}_{x \backsim M} p(x) \\ \text{s.t.} \quad & \mathop{\mathbb{E}}_{x \backsim M} q_i(x) \leq b_i \,, \qquad \forall i \in [m] \,. \end{aligned}$$

We may allow degree-1 terms in the polynomials p and q_i by using a dummy variable. Simply add the variable x_0 with the constraints $\mathbb{E}_{x \backsim M} x_0^2 \leq 1$ and $\mathbb{E}_{x \backsim M} x_0^2 \geq 1$. Then since nothing in the SDP distinguishes x_0 from actually being equal to 1, we may treat it as 1, so that $\mathbb{E}_{x \backsim M} x_i := \mathbb{E}_{x \backsim M} x_0 x_i$.

4.1 Solving an SDP

Recall from a previous lecture that the ellipsoid algorithm only required access to a separating-hyperplane oracle to find a feasible point of a convex body.

The degree-2 homogeneous polynomial constraints and objective functions in x, are linear in the actual variables $\mathbb{E}_{x \backsim M} x_i x_j$ of the semidefinite program. Thus they serve as their own separating-hyperplane oracles, like for LPs. As there are m such constraints each referring to up to n^2 variables, checking each of these takes $O(mn^2)$ time.

The positive-semidefinite-ness constraint $M \succeq 0$ can be characterized as an infinite family of linear constraints: $v^T M v \geq 0$ for all $v \in \mathbb{R}^n$. Then simply taking an eigendecomposition of M finds you some v for which the constraint is violated, if it exists. An eigendecomposition can be done in $O(n^3)$ arithmetic operations.

All-in-all, an iteration of ellipsoid takes $O(n^2 m + n^3)$ arithmetic operations to do. As each iteration cuts the size of the search space by a multiplicative factor, it takes $O((n^2 m + n^3) \log(\varepsilon^{-1}))$ arithmetic operations to find a feasible point within $\varepsilon^{1/n}$

additive error of the objective function, assuming that the feasible space has volume at least ε, and that the starting ellipsoid has constant volume.

A word of warning: this running time depends on the specific formulation of semidefinite programming being used.

5 Goemans-Williamson algorithm for MAX-CUT via semidefinite programming

Recall that earlier in this lecture, we formulated MAX-CUT as the following mathematical program:

$$\begin{aligned} \max_{x} \quad & \sum_{\{u,v\} \in E} \tfrac{1}{2}(1 - x_u x_v) \\ \text{s.t.} \quad & x_v \in \{-1, 1\}\,, \qquad \forall v \in V\,. \end{aligned}$$

To make this a semidefinite programming relaxation, we simply insert expectation operators:

$$\begin{aligned} \max_{M \succeq 0} \quad & \mathop{\mathbb{E}}_{x \backsim M} \sum_{\{u,v\} \in E} \tfrac{1}{2}(1 - x_u x_v) \\ \text{s.t.} \quad & \mathop{\mathbb{E}}_{x \backsim M} x_v^2 = 1\,, \qquad \forall v \in V\,. \end{aligned}$$

We may solve this SDP to obtain a covariance matrix M maximizing the value $c(M) = \mathbb{E}_{x \backsim M} \sum_{\{u,v\} \in E} \frac{1}{2}(1 - x_u x_v)$. Since this SDP generalizes actual solutions to MAX-CUT (take M to be the covariance of a constant distribution over a solution), this value is at least as large as the capacity of the actual maximum cut.

We *round* the SDP solution M (akin to "rounding" a fractional LP solution) by sampling a random point z from the Gaussian distribution $\mathcal{N}(0, M)$ with mean 0 and covariance M. Then taking the cut to be $A_z = \{v \in V : z_v > 0\}$ reduces the value from the SDP by at worst a factor of 0.878 (rounded down) in expectation, as explained in the next section, giving us an 0.878-approximation to MAX-CUT.

5.1 Analysis

We start by using linearity of expectation to compute

$$\begin{aligned}\mathop{\mathbb{E}}_{z\sim\mathcal{N}(0,M)} c(A_z) &= \mathop{\mathbb{E}}_{z\sim\mathcal{N}(0,M)} \sum_{\{u,v\}\in E} \mathbb{1}\left[z_u > 0 \oplus z_v > 0\right] \\ &= \sum_{\{u,v\}\in E} \mathop{\mathbb{E}}_{z\sim\mathcal{N}(0,M)} \mathbb{1}\left[z_u > 0 \oplus z_v > 0\right] \\ &= \sum_{\{u,v\}\in E} \Pr_{z\sim\mathcal{N}(0,M)}\left[z_u > 0 \oplus z_v > 0\right] .\end{aligned}$$

Fix some $\{u, v\} \in E$ and we'll evaluate $p := \Pr_{z\sim\mathcal{N}(0,M)}[z_u > 0 \oplus z_v > 0]$.

Note that since p depends only on z_u and z_v, we may restrict our attention to the 2-dimensional case, with the covariance matrix

$$M' = \begin{pmatrix} 1 & \mathop{\mathbb{E}}_{x\backsim M} x_u x_v \\ \mathop{\mathbb{E}}_{x\backsim M} x_u x_v & 1 \end{pmatrix},$$

so that

$$p = \Pr_{z\sim\mathcal{N}(0,M')}\left[z_1 > 0 \oplus z_2 > 0\right] .$$

Recalling that $z \sim \mathcal{N}(0, M')$ has the same distribution as $M'^{1/2}y$ when y is sampled from a standard Gaussian $y \sim \mathcal{N}(0, \mathrm{Id})$,

$$p = \Pr_{y\sim\mathcal{N}(0,\mathrm{Id})}\left[(M'^{1/2}y)_1 > 0 \oplus (M'^{1/2}y)_2 > 0\right] .$$

Equivalently, letting μ be the first column of $M'^{1/2}$ and ν be the second column,

$$p = \Pr_{y\sim\mathcal{N}(0,\mathrm{Id})}\left[\langle\mu, y\rangle > 0 \oplus \langle\nu, y\rangle > 0\right] .$$

Fact 5.1. *The standard multivariate Gaussian distribution* $\mathcal{N}(0, \mathrm{Id})$ *is* rotationally invariant.

Notice that the test $\langle\mu, y\rangle > 0$ simply selects a half-plane through the origin with normal vector μ. Since $y \sim \mathcal{N}(0, \mathrm{Id})$ is rotationally invariant, this is equivalent

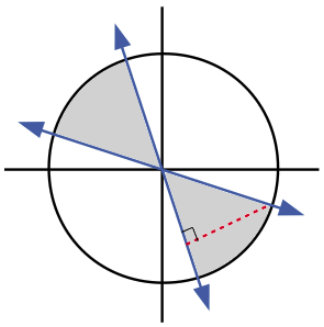

Figure 1: The area (in gray) of a circle in the symmetric difference of two half-spaces. The dot product (dotted red line) of the vectors defining those two half-spaces is the cosine of the angle between them.

to asking how much of the area of a circle is in that symmetric difference of two half-planes (Figure 1). Let a and b be so that $(a \;\; b)^T = \mu$ and $(b \;\; a)^T = \nu$. Since

$$\begin{pmatrix} 1 & \mathop{\mathbb{E}}_{x \smile M} x_u x_v \\ \mathop{\mathbb{E}}_{x \smile M} x_u x_v & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}^2 = \begin{pmatrix} a^2 + b^2 & 2ab \\ 2ab & a^2 + b^2 \end{pmatrix},$$

we can conclude that both $\mu = (a \;\; b)^T$ and $\nu = (b \;\; a)^T$ are unit vectors, and that their dot product is $\mathbb{E}_{x \smile M} x_u x_v$. This means that the angle between the two vectors is given by $\arccos(\mathbb{E}_{x \smile M} x_u x_v)$, so that

$$p = \frac{1}{\pi} \arccos\left(\mathop{\mathbb{E}}_{x \smile M} x_u x_v \right).$$

And it just so happens that $\frac{1}{\pi}\arccos(x) \geq 0.878 \cdot \frac{1}{2}(1 - x)$ for all $x \in [-1, 1]$. So therefore,

$$p \geq 0.878 \cdot \tfrac{1}{2}\left(1 - \mathop{\mathbb{E}}_{x \smile M} x_u x_v\right),$$

so that

$$\begin{aligned} \mathop{\mathbb{E}}_{z \sim \mathcal{N}(0,M)} c(A_z) &\geq \sum_{\{u,v\} \in E} 0.878 \cdot \tfrac{1}{2}\left(1 - \mathop{\mathbb{E}}_{x \smile M} x_u x_v\right) \\ &= 0.878 \mathop{\mathbb{E}}_{x \smile M} \sum_{\{u,v\} \in E} \tfrac{1}{2}(1 - x_u x_v) \\ &= 0.878 \, c(M) \\ &\geq 0.878 \max_{U \subseteq V} c(U). \end{aligned}$$

6 Relationship with usual presentation of SDP

Recall our generic SDP, with degree-2 homogeneous polynomials $p(x)$ and $q_i(x)$ and $b \in \mathbb{R}^m$,

$$\begin{aligned} \min_{M \succeq 0} \quad & \mathop{\mathbb{E}}_{x \backsim M} p(x) \\ \text{s.t.} \quad & \mathop{\mathbb{E}}_{x \backsim M} q_i(x) \leq b_i \,, \qquad \forall i \in [m] \,. \end{aligned}$$

Proposition 6.1. *$M \succeq 0$ if and only if $M \in \mathbb{R}^{n \times n}$ is a* Gram matrix *of n-dimensional vectors. That is, if and only if there is a sequence of vectors $z_1, \ldots z_n \in \mathbb{R}^n$ such that $M_{i,j} = \langle z_i, z_j \rangle$.*

Proof. By Proposition 3.2, $M \succeq 0$ if and only if $M = AA^T$ for some matrix A. In fact, we may take $A = M^{1/2}$. Let z_i be the ith row of A. Then $M_{i,j} = e_i^T AA^T e_j = z_i^T z_j = \langle z_i, z_j \rangle$. □

Therefore, $\mathbb{E}_{x \backsim M} x_i x_j$ may be interpreted as $\langle z_i, z_j \rangle$ for some set of vectors $z_i, \ldots, z_j \in \mathbb{R}^n$.

This gives us precisely the usual presentation of an SDP, which is

$$\begin{aligned} \min_{z_1, \ldots, z_n \in \mathbb{R}^n} \quad & \sum_{i,j \in [n]} c_{i,j} \langle z_i, z_j \rangle \\ \text{s.t.} \quad & \sum_{i,j \in [n]} a_{i,j,k} \langle z_i, z_j \rangle \leq b_k \,, \qquad \forall k \in [m] \,. \end{aligned}$$

7 Duality of SDPs and coefficient matrices of polynomials

Consider again the generic SDP, with degree-2 homogeneous polynomials $p(x)$ and $q_i(x)$ and $b \in \mathbb{R}^m$,

$$\begin{aligned} \min_{M \succeq 0} \quad & \mathop{\mathbb{E}}_{x \backsim M} p(x) \\ \text{s.t.} \quad & \mathop{\mathbb{E}}_{x \backsim M} q_i(x) \leq b_i \,, \qquad \forall i \in [m] \,. \end{aligned}$$

The dual of this SDP is the following:

$$\begin{aligned} \max_{y \in \mathbb{R}^m} \quad & \langle b, y \rangle \\ \text{s.t.} \quad & \sum_{i \in [m]} y_i \, q_i(x) \leq p(x) \,, \qquad \forall x \in \mathbb{R}^n \,. \end{aligned}$$

By Stengle's Positivstellensatz, that last constraint holds for all $x \in \mathbb{R}^n$ precisely when $p(x) - \sum_{i \in [m]} y_i \, q_i(x)$ can be expressed as a sum of square polynomials.

This may in fact be expressed as another PSD-ness contraint, if we represent these polynomials as *coefficient matrices*.

Definition 7.1. *For a homogeneous degree-2 polynomial $p : \mathbb{R}^n \to \mathbb{R}$ with $p(x) = \sum_{i,j} c_{i,j} x_i x_j$, let its* coefficient matrix *be $C : \mathbb{R}^{n \times n}$ with $C_{i,j} = \frac{1}{2}(c_{i,j} + c_{j,i})$.*

Then

$$\mathop{\mathbb{E}}_{x \backsim M} p(x) = \langle C, M \rangle := \mathrm{Tr}(CM) = \sum_{i,j} C_{i,j} M_{i,j} \,.$$

Observe that every square polynomial has a rank-1 positive-semidefinite coefficient matrix: if $p(x) = \langle v, x \rangle^2$ with $\langle v, x \rangle$ being a generic linear function, then the coefficient matrix of p is equal to vv^T. Therefore, every sum of square polynomials has a PSD coefficient matrix. And conversely, since every PSD matrix can be written as a sum of rank-1 PSD terms, every PSD coefficient matrix is a sum of square polynomials.

Weak duality: The value of the dual, when it exists, is at most the value of the primal: $\mathbb{E}_{x \backsim M^*} p(x) \geq \langle b, y^* \rangle$ where M^* is an optimal primal solution and y^* is an optimal dual solution. This is seen from the following derivation:

$$\begin{aligned} \mathop{\mathbb{E}}_{x \backsim M^*} p(x) - \langle b, y^* \rangle &= \mathop{\mathbb{E}}_{x \backsim M^*} p(x) - \sum_{i \in [m]} b_i y_i^* \\ &\geq \mathop{\mathbb{E}}_{x \backsim M^*} p(x) - \sum_{i \in [m]} \mathop{\mathbb{E}}_{x \backsim M^*} q_i(x) y_i^* \\ &= \mathop{\mathbb{E}}_{x \backsim M^*} \left(p(x) - \sum_{i \in [m]} q_i(x) y_i^* \right) \,. \end{aligned}$$

Recall that $p(x) - \sum_{i \in [m]} q_i(x) y_i^*$ is non-negative for all x, so its expectation must also be non-negative.

Strong duality: Not all SDPs satisfy strong duality. But they do under *Slater's condition*, which says that there is a feasible solution M_0 to the primal such that $M_0 \succ 0$ is strictly positive-definite (equivalently, PSD and full rank). When that condition is satisfied, then $\mathbb{E}_{x \backsim M^*} p(x) = \langle b, y^* \rangle$.

8 Complexity theory and the hardness of approximation

We saw that Goemans-Williamson's algorithm gets a 0.878-approximation for MAX-CUT. We might ask if there's a way to do better than that.

This gets us into the topic of the Hardness of Approximation. There have been two main themes in approximation algorithms for maximum-constraint-satisfaction problems: most of the time, either we can show that it's NP-hard to achieve any approximation ratio better than the one attained by a random assignment (such as for MAX-3-SAT or MAX-3-XOR), or we know an algorithm that does better than random, but we don't know if it's possible to do even better than that, or if doing any better would be NP-hard. MAX-CUT is among the latter: we don't know if it's NP-hard to achieve an α-approximate MAX-CUT for any α strictly between 0.879 and 1.

The Unique Games Conjecture is a central conjecture in Hardness of Approximation, which, if true, would imply hardness results exactly matching many of the known algorithms in the latter case, including MAX-CUT. The conjecture roughly posits that it is NP-hard to achieve any constant-factor approximation to the UNIQUE-LABEL-COVER problem: finding a labeling of the vertices of a graph that maximizes the number of satisfied edge constraints, when the edge constraints are guaranteed to satisfy the property that for any label assigned to one of its incident vertices, there is only one label possible for the other vertex that would satisfy the constraint.

There is a reduction by Khot, Kindler, Mossel, and O'Donnell in 2005, from a constant-factor approximation of UNIQUE-LABEL-COVER to a 0.879-approximation

of MAX-CUT. Thus if the Unique Games Conjecture is true, the Goemans-Williamson algorithm achieves the optimal polynomial-time approximation assuming P $\neq$ NP.

To give a rough sense of the reduction: it relies on the "Majority is Stablest" theorem, which is a statement about the geometry of the hypercube, roughly saying that planar cuts through the origin minimize the ratio of edges cut to the number of vertices in the smaller half of the cut. The reduction takes an instance of UNIQUE-LABEL-COVER with vertex set V and L the set of possible labels for each vertex, and creates an instance of MAX-CUT with vertex set $V \times 2^L$ (so, ordered pairs consisting of a vertex in V along with a subset of L). The idea here is be that a label assignment $x \in L^V$ in the label cover instance will correspond to the cut in the MAX-CUT instance given by $A = \{(u, S) \in V \times 2^L : x_u \in S\}$. The edges in the MAX-CUT instance are a bit complicated to specify, but they basically are set up to have a large number of edges cut when (1) label-cover constraint are satisfied (2) the implied labelling of each vertex is consistent—i.e. for each u there is an ℓ such that $(u, S) \in A \iff \ell \in S$. Majority is Stablest comes in to ensure (2): it tells us that the worst non-consistent labellings will pass the consistency check with $1 - 0.879$ probability as $|L| \to \infty$.

This construction is an example of a *probabilistically checkable proof*: so-called because an assignment to the MAX-CUT instance that achieves good value serves as a proof that the UNIQUE-LABEL-COVER instance also has good value: furthermore it's a proof that can be checked probabilistically *by only inspecting a few bits.* If (and only if) the UNIQUE-LABEL-COVER instance indeed has a good labelling, then a good cut exists for the MAX-CUT instance. And if (and only if) that good cut exists, then you only need to check a few randomly selected edges to be convinced that it does indeed achieve good value.

All known hardness-of-approximation results to date follow this sort of construction.