Suggested reading See Section 6.2 (pages 143-147 in electronic version) of Williamson-Shmoys book "Design of approximation algorithms" for an overview of the Geomans-Williamson Max-Cut algorithm. The Cheeger-Alon-Milman Inequality is covered in many places. One good source is Trevisan's CS359G Lecture Notes `http://theory.stanford.edu/~trevisan/cs359g/`, Lectures 3 and 4.

A tale of two problems

Setting $G = (V, E)$ is d-regular graph on n vertices. It has normalized adjacency matrix A ($A_{i,j} = 1/d$ if $(i, j) \in E$ and $A_{i,j} = 0$ otherwise). Let $L = I - A$ be its normalized *Laplacian matrix.*

(Uniform) Sparsest cut problem Find set $S \subseteq V$ that minimizes $\phi(S) = nE(S, \overline{S})/(d|S||\overline{S}|)$. Let $\phi(G) = \min_S \phi(S)$.

Max cut (min uncut) problem Find set $S \subseteq V$ that maximizes $\mathrm{cut}(S) = E(S, \overline{S})/|E|$, or equivalently minimizes $\mathrm{uncut}(S) = 1 - \mathrm{cut}(S)$. Let $\mathrm{cut}(G) = \max_S \mathrm{cut}(S)$, $\mathrm{uncut}(G) = 1 - \mathrm{cut}(G)$.

Results A random subset of measure 1/2 will cut half the edges, and in particular this gives an algorithm achieving a cut of value at least $\mathrm{cut}(G)/2$ for the Max cut problem, or achieving value at most $1 - (1 - \phi(G))/2$ for the sparsest cut problem. In fact, this algorithm for Max Cut was suggested by Erdös in 1967, and is one of the first analyses of any approximation algorithm.

A priori, it is not so clear how to beat this. Lets consider the case of Max Cut. In a random d-regular graph (which is an excellent expander), one cannot cut more than a $1/2 + \epsilon$ fraction of the edges (where the ϵ goes to zero at least with d (maybe with n?)). But locally, it is hard to distinguish a random d-regular graph from a random d-regular "almost bipartite" graph, where we split the graph into two equal parts L and R and each edge is with probability ϵ inside one of those parts and with probability $1 - \epsilon$ between them. Such a graph G obviously has $\mathrm{cut}(G) \geq 1 - \epsilon$ but every neighborhood of it looks like a d-regular tree, just as in the case of a random d-regular graph. For this reason, "combinatorial" algorithms have a hard time getting an approximation factor better than 1/2 for Max Cut.

- $\exists$ poly-time algorithm to find S with $\phi(S) = O(\sqrt{\phi(G)})$. (Cheeger, Alon-Milan)
- $\exists$ poly-time algorithm to find S with $\phi(S) = O(\sqrt{\log n}\phi(G))$ (Arora-Rao-Vazirani)
- $\exists$ poly-time algorithm to find S with $\mathrm{uncut}(S) = O(\sqrt{\mathrm{uncut}(G)})$ (Goemans-Williamson.) (Corollary: there is an algorithm to find S with $\mathrm{cut}(S) \geq \alpha \mathrm{cut}(G)$ for some $\alpha > 1/2$; value of α turns out to be 0.878.. (?).)
- $\exists$ poly-time algorithm to find S with $\mathrm{uncut}(S) = O(\sqrt{\log n}\mathrm{uncut}(G))$ (Agarwal-Charikar-Makarychev-Makarychev, following ARV.)
- Assuming Unique Games Conjecture, there is no poly-time algorithm to find S with $\mathrm{uncut}(S) = o(\sqrt{\mathrm{uncut}(G)})$.
- Assuming Small-Set Expansion conjecture, there is no poly-time algorithm to find S with $\phi(S) = o(\sqrt{\phi(G)})$.

Linear algebra view Let x be the $\{\pm 1\}$ characteristic vector of S (i.e., $x_i = +1$ if $i \in S$ and $x_i = -1$ otherwise).

Then

$$\langle x, Lx \rangle = \sum_i x_i^2 - \tfrac{1}{d} \sum_{i \sim j} x_i x_j = \tfrac{1}{2d} \sum_{i \sim j} (x_i - x_j)^2 = 4E(S, \overline{S})/d$$

If x is the mean-zero characteristic vector (i.e., $x_i = +|\overline{S}|$ if $i \in S$ and $x_i = -|S|$ if $i \in \overline{S}$) then

$$\langle x, Lx \rangle = \tfrac{1}{2d} \sum_{i \sim j} (x_i - x_j)^2 = (|S| + |\overline{S}|)^2 E(S, \overline{S})/d = n^2 E(S, \overline{S}) \ .$$

Note that in this case $\|x\|^2 = |S||\overline{S}|^2 + |\overline{S}||S|^2 = n|S||\overline{S}|$

Max cut The heart of the Goemans-Williamson theorem is the following:

$\{\pm 1\}$ quadratic sampling lemma: Let $\{x\}$ be a degree 2 pseudo-distribution satisfying the constraints $\{x_i^2 = 1\}$. Then there exists an actual distribution $\{z\}$ over $\{\pm 1\}^n$ such that for every i, j, if $\tilde{\mathbb{E}} x_i x_j = 1 - \epsilon$, then $\mathbb{E} z_i z_j \leq -1 + O(\sqrt{\epsilon})$.

Proof By shifting we can assume that $\tilde{\mathbb{E}} x_i = 0$ for all i (we can also scale it so that $\tilde{\mathbb{E}} x_i^2 = 1$ though it doesn't matter for this proof). Since $\{x\}$ is a pseudo-distribution, its second moment matrix $M = \tilde{\mathbb{E}} x^{\otimes 2} = \tilde{\mathbb{E}} x x^\top$ is psd. Hence, we can write $M = V^\top V$ where V is a $d \times n$ matrix with columns $v_1, \ldots, v_n$ and so $M_{i,j} = \langle v_i, v_j \rangle$. Choose a random standard Gaussian vector $g = (g_1, \ldots, g_n)$ and let $z_i = \langle v_i, g \rangle$.

Then, for every i, j, we get that

$$\mathbb{E} z_i z_j = \mathbb{E} \langle v_i, g \rangle \langle v_j, g \rangle = \sum_{a,b} v_i(a) g_a v_i(b) g_b = \sum a_i(a) v_j(a) = \langle v_i, v_j \rangle = M_{i,j}$$

using the fact that the Gaussians are independent and so $\mathbb{E} g_a g_b$ equals 0 if $a \neq b$ and equals 1 otherwise.

Corollary The distribution $\{z\}$ satisfies

$$\mathbb{E} \langle z, Lz \rangle = \mathbb{E} 4n \cdot \mathrm{cut}(z) \geq (4 - O(\sqrt{\epsilon}))n$$

(Proof: exercise - uses concavity of $\sqrt{\cdot}$ function.)

Proof of Lemma The construction of the distribution $\{z\}$ is that we simply use $z = \mathrm{sign}(y)$ where $\{y\}$ is the Gaussian distribution matching the first two moments of $\{x\}$. We thus simply need to prove the following lemma: (can you see why)

Lemma Let Y, Y' be two Gaussian random variables satisfying $\mathbb{E} Y^2 = \mathbb{E} Y'^2 = 1$ and $\mathbb{E}(Y - Y')^2 \geq 4 - \epsilon$. Then $\mathbb{E}(\mathrm{sign}(Y) - \mathrm{sign}(Y'))^2 \geq 4 - O(\sqrt{\epsilon})$.

Proof of lemma Condition is equivalent to $\mathbb{E} Y Y' \leq -1 + O(\epsilon)$, and so we simply have to compute what's the probability that two Gaussians that almost perfectly anti-correlated (correlation $-1 + \epsilon$) agree in their signs, and it is not too hard to see that it is $O(\sqrt{\epsilon})$. (Exercise)

Corollary There is a polynomial time algorithm finding S with $\mathrm{uncut}(S) = O(\sqrt{\mathrm{uncut}(G)})$.

Sparsest cut We consider the case that there is a pseudo-distribution $\{T\}$ over sets T of size exactly $n/2$ such that $|E(T, \overline{T})| \leq \epsilon dn$. We define the pseudo-distribution $\{x\}$ so that $x_i = +1$ if $i \in S$ and $x_i = -1$ otherwise.

Now let $\{z\}$ be the distribution obtained by the 0/1 quadratic sampling lemma. Then for every x_i, x_j, $\mathbb{E}1 - z_i z_j \leq O(\sqrt{\mathbb{E}1 - x_i x_j})$ and so, letting $S = \{i : z_i = +1\}$,

$$4\mathbb{E}E(T, \overline{T}) = d\langle z, Lz\rangle \leq O(\sqrt{\epsilon}d)\|z\|^2 = O(\sqrt{\epsilon}dn)$$

The condition that $\{T\}$ satisfies that the sets have size exactly $n/2$ means that $\{x\}$ satisfies the constraint $\{\sum x_i = 0\}$ which means in particular that

$$\mathbb{E}(\sum y_i)^2 = \tilde{\mathbb{E}}(\sum x_i)^2 = 0$$

which can be used to show that $\Pr[|S| \leq 0.9n] \geq 0.01$, which suffices for the result (exercise).

The exercises discuss the idea needed to generalize the result when the pseudo-distribution $\{T\}$ is over sets of size $k \leq n/2$.

In fact, the conditions we posit do not require the pseudo-distribution to be over 0/1 or ± 1 valued vectors, and hence (after shifting) it is enough to find a pseudo (or actual) distribution over vectors $x \in R^n$ such that $\sum x_i = 0$ and $x^\top Lx \leq \phi\|x\|^2$. This leads to the more usual description of Cheeger's inequality as relating expansion to the second largest eigenvalue of the adjacency matrix (or equivalently, the second smallest eigenvalue of the Laplacian).