SDPs, Max Cut, and Goemans Williamson[1]

- In this lecture, we look at "SDP rounding" for the Max Cut problem. Let's recall a few things from the last lecture. In the max cut problem, one is given an undirected graph $G = (V, E)$ and every edge e has a non-negative weight $w(e)$. The objective is to find a subset $S \subseteq V$ such that $w(\partial S) = \sum_{e \in \partial S} w(e)$ is maximized. The SDP relaxation for the maximum cut problem is the following.

$$\texttt{opt} \leq \textsf{sdp} := \max \quad \frac{1}{2} \cdot \sum_{(u,v) \in E} w(u,v) \cdot (1 - \mathbf{X}_{uv}) \tag{Max Cut SDP}$$

$$\mathbf{X}_{vv} = 1, \quad \forall v \in V \tag{1}$$

$$\mathbf{X} \succcurlyeq 0, \tag{2}$$

  Next, we decompose the SDP solution X to obtain n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which live in $\mathbb{R}^d$ such that $\mathbf{X}_{ij} = \mathbf{v}_i^\top \mathbf{v}_j$; here d is the rank of X and is at most n. In particular, $\|\mathbf{v}_i\| = X_{ii} = 1$, that is, the vectors are unit vectors (in ℓ_2-norm). The SDP can therefore be recast as

$$\texttt{opt} \leq \textsf{sdp} = \frac{1}{2} \sum_{(i,j) \in E} w(i,j) \cdot \left(1 - \mathbf{v}_i^\top \mathbf{v}_j\right) \quad : \quad \|\mathbf{v}_i\|_2 = 1, \forall i, \quad \mathbf{v}_i \in \mathbb{R}^d \tag{3}$$

  The objective is to take these n high-dimensional vectors, one corresponding to each vertex $i \in V$, and then somehow obtain a cut $S \subseteq V$ whose value can be compared with $\textsf{sdp}$. This procedure is often called SDP "rounding", although rounding is not quite what is occurring here.

- *SDP "Integrality" Gap.* Let us discuss the notion of "integrality gap" of the SDP relaxation. This is akin to the notion of integrality gap for LPs, and is defined as

$$\alpha_{\text{(Max Cut SDP)}} := \inf_{\mathcal{I} \,:\, \text{Max Cut Instance}} \frac{\texttt{opt}(\mathcal{I})}{\textsf{sdp}(\mathcal{I})}$$

  The best approximation one can hope to prove using an SDP relaxation is this quantity and we would like to prove this as large as possible. However, any example would provide an upper bound on this. Before we describe the algorithm, let us give a simple example giving an upper bound $\alpha_{\text{(Max Cut SDP)}}$; this perhaps helps in getting a feel of the algorithm.

- *A Simple Integrality Gap Example.* The example is just the 5-cycle C_5. The maximum cut value is 4 obtained by taking any two non-consecutive vertices on one side. We now describe a SDP solution to (3) which has a higher value than 4. We need to show 5 unit vectors corresponding to the 5 vertices such that for neighboring vectors, $\mathbf{v}$ and $\mathbf{v}'$ say, the dot-product $\mathbf{v}^\top \mathbf{v}'$ is as small as possible. Or geometrically, they should be making as ***obtuse*** an angle with each other. Can you see how to do this? See Figure 1.

  Note that the vectors in the example above lay in 2-dimension. One may think that perhaps the sdp value for this example can be increased by considering vectors in higher dimensions. But turns out

[1] *Lecture notes by Deeparnab Chakrabarty. Last modified : 3rd Mar, 2022*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

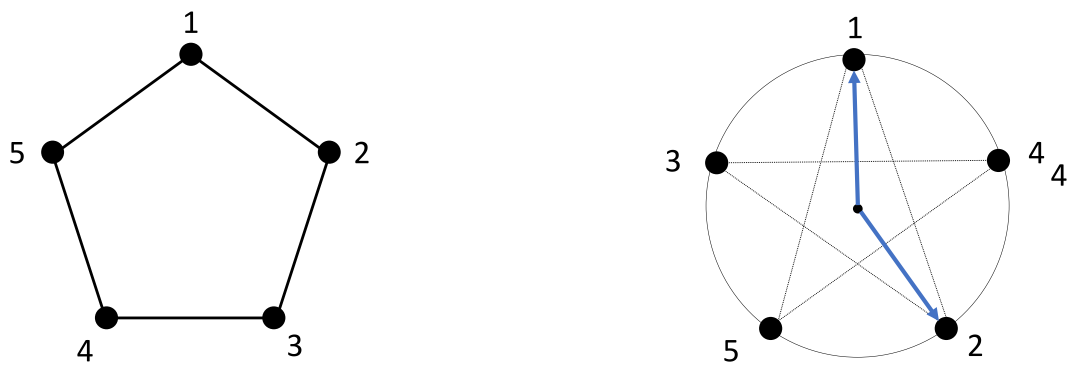

Figure 1: *The C_5 embedded onto the boundary of a unit circle. The angle between two neighboring vectors is $4\pi/5 = 144°$ and thus each edge contributes $1 - \cos(4\pi/5) \approx 1.809$, and the* sdp *solution is ≈ 4.5225 implying $\alpha_{\text{(Max Cut SDP)}} \leq 0.8845$.*

that this is the optimal SDP solution (one can check this by plugging into an SDP solver). One may also imagine that there may be much more complicated examples proving much better upper bounds on $\alpha_{\text{(Max Cut SDP)}}$, and perhaps this value could be closer to 0.5, which is the integrality gap of the "natural LP relaxation" for the max-cut problem. It was indeed a surprise that the true integrality gap is quite close to 0.88. And the algorithm is *extremely* simple and is one of the most famous algorithms in approximation algorithms.

- ***The Goemans-Williamson Algorithm.*** Let's recall again what we have : we have n vectors embedded on the boundary of a high-dimensional sphere. The embedding is such that vectors corresponding to end points of an edge have a "repulsive force" and are trying to be on antipodal points of the sphere. Our goal is to find a partition of these vertices into two parts. The Goemans-Williamson algorithm, in this author's imagination, is as follows. This high-dimensional sphere is an orange, which we give a "random spin" and slice it through the center with a high-dimensional samurai sword. For any pair of neighboring vertices, if they are "far away" on the orange, then it's "more likely" they will be sundered by the slice. More precisely, we choose a ***random unit vector*** $\mathbf{r} \in \mathbb{R}^d$ and cut the high-dimensional ball using the hyperplane for which this $\mathbf{r}$ is the normal. The partition $S \subseteq V$ is formed by picking the vectors which fall in one side of this hyperplane. Formally,

procedure GOEMANS-WILLIAMSON($\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^d$, $\|\mathbf{v}_i\|_2 = 1$):
    Sample $\mathbf{r} \in \mathbb{R}^d$ with $\|\mathbf{r}\|_2 = 1$ uniformly. ▷ *How does one do that?*
    $S = \{i : \mathbf{v}_i^\top \mathbf{r} \geq 0\}$.
    return $(S, V \setminus S)$.

Theorem 1. The expected weight of the cut returned by GOEMANS-WILLIAMSON is $\alpha_{\mathsf{GW}} = \frac{2}{\pi} \cdot \inf_{z \in [-1,1]} \frac{\arccos z}{1-z} \approx 0.87856$.

Proof. Fix an edge $(i,j) \in E$. We prove that $\mathbf{Pr}[(i,j) \in \partial S] \geq \alpha_{\mathsf{GW}} \cdot \left(\frac{1}{2} \cdot (1 - \mathbf{v}_i^\top \mathbf{v}_j)\right)$; the proof then follows via linearity of expectation.

Now, consider the 2-dimensional circle C on the plane P defined by $\mathbf{v}_i, \mathbf{v}_j$ and $\mathbf{0}$. Consider the hyperplane $H := \{\mathbf{x} : \mathbf{x}^\top \mathbf{r} = 0\}$ and let $\mathbf{h} := H \cap P$ be the line passing through the centre of C. The first observation is that $(i,j) \in \partial S$ if and only if $\mathbf{v}_i$ and $\mathbf{v}_j$ lie on opposite sides of the "diameter" $\mathbf{h}$. The second key observation is that if $\mathbf{r}$ is uniformly sampled on the d-dimensional sphere, then by symmetry on any great circle (whose centre is $\mathbf{0}$) the projection of $\mathbf{r}$ on that plane is uniformly random on that circle. In other words, $\mathbf{h}$ is uniformly distributed among all diameters. See Figure 2 for an illustration.

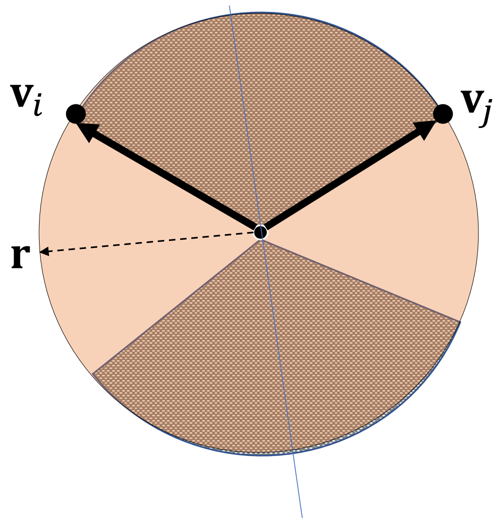

Figure 2: *The orange restricted to the plane containing $\mathbf{v}_i$, $\mathbf{v}_j$ and the centre. The blue $\mathbf{h}$ is the intersection of the $(d-1)$ dimensional samurai sword and this plane. Any diameter is equally likely, and the slice split i and j if and only if it's in the shaded region. The angle is $\arccos(\mathbf{v}_i^\top \mathbf{v}_j)$ and there are 2 of them. This divided by 2π is the probability i and j are split.*

Therefore if we say θ is the angle between $\mathbf{v}_i$ and $\mathbf{v}_j$, which is the angle subtended on this plane P and equals $\arccos\left(\mathbf{v}_i^\top \mathbf{v}_j\right)$, then the probability i and j are separated is $\frac{\theta}{\pi}$. In other words,

$$\mathbf{Pr}[(i,j) \in \partial S] = \frac{\arccos\left(\mathbf{v}_i^\top \mathbf{v}_j\right)}{\pi} = \underbrace{\left(\frac{2}{\pi} \cdot \frac{\arccos\left(\mathbf{v}_i^\top \mathbf{v}_j\right)}{1 - \mathbf{v}_i^\top \mathbf{v}_j}\right)}_{\geq \alpha_{\mathsf{GW}} \text{ by definition}} \cdot \left(\frac{1}{2} \cdot (1 - \mathbf{v}_i^\top \mathbf{v}_j)\right) \qquad \square$$

- *The Goemans-Williamson Constant.* The function $g(z) := \frac{2}{\pi}\frac{\arccos z}{1-z}$ obtains its minimum value at $z \approx -0.69$. This can be seen in Figure 3. That is, when the angle between the endpoints of the edges is roughly $133.63°$. This minimum value of $g(z)$ is called α_{GW}, the Goemans-Williamson constant.

  It is instructive to ask what occurs if the angle between the endpoints of the edges is much larger, and really close to $180°$, or π-radians. More precisely, if the contribution of the edge (i,j) to the SDP value $\geq (1-\varepsilon)$, then what's the probability this edge would be cut by the GOEMANS-WILLIAMSON algorithm? As can be seen from the plot in Figure 3, as $z \to \cos(\pi) = -1$, the value $g(z) \to 1$ as well. A non-asymptotic statement is also possible as stated below.

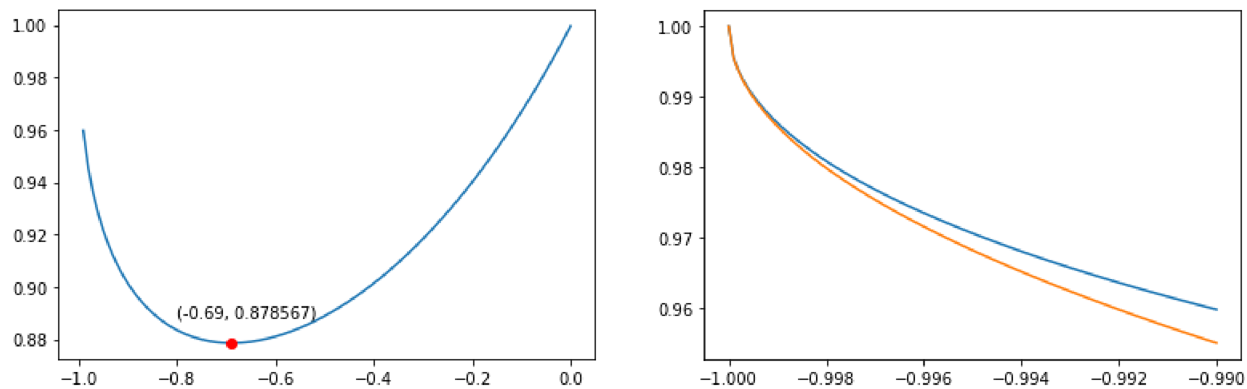

Figure 3: *The left figure shows a plot of* $g(z) := \frac{2}{\pi} \cdot \frac{\arccos z}{1-z}$ *in* $z = [-1, 0]$ *marking the minimum point as a red dot. This plot is a proxy of how the approximation ratio behaves if all the* $\mathbf{v}_i^\top \mathbf{v}_j$*'s were* z*. The right figure shows the same plot in the range* $z = [-1, -0.99]$*. It is overlaid with the plot* $h(z) = 1 - \frac{2}{\pi} \cdot \sqrt{\frac{1+z}{2}}$ *in orange. Note if* $\frac{1-z}{2} = 1 - \varepsilon$*, then* $H(z)$ *is plotting* $1 - \frac{2}{\pi}\sqrt{\varepsilon}$*.*

Lemma 1. If $\frac{1}{2} \cdot \left(1 - \mathbf{v}_i^\top \mathbf{v}_j\right) \geq (1 - \varepsilon)$, then $\mathbf{Pr}[(i,j) \in \partial S] \geq 1 - \sqrt{2\varepsilon}$.

Proof. We may assume ε is small enough, in particular $\leq 1/2$, for otherwise the RHS holds vacuously. Let's denote the angle between $\mathbf{v}_i$ and $\mathbf{v}_j$ as $\pi - \theta$. Then, $\mathbf{Pr}[(i,j) \in \partial S] = 1 - \frac{\theta}{\pi}$.

Since $\varepsilon \leq 0.5$, we get that $\theta \leq \frac{\pi}{2} < 2$. And therefore, from the Taylor expansion of $\cos(x)$ around π. More precisely,

$$\begin{aligned} \cos(\pi - \theta) &= \cos(\pi) - \frac{\cos(\pi)\theta^2}{2!} + \frac{\cos(\pi)\theta^4}{4!} \cdots \\ &= -1 + \frac{\theta^2}{2!} - \frac{\theta^4}{4!} + \frac{\theta^6}{6!} - \frac{\theta^8}{8!} \cdots \qquad > -1 + \frac{\theta^2}{3} \end{aligned} \tag{4}$$

for $\theta < 2$. To be clear, this is because for such small θ, $\frac{\theta^2}{6} > \frac{\theta^4}{24}$, and $\frac{\theta^{2k}}{(2k)!} > \frac{\theta^{2k+2}}{(2k+2)!}$ for $k \geq 3$.

Now, $\cos(\pi - \theta) = \mathbf{v}_i^\top \mathbf{v}_j \leq -1 + 2\varepsilon$ by the premise of the lemma. Therefore, using (4), we get $\theta < \sqrt{6\varepsilon}$, implying $\mathbf{Pr}[(i,j) \in \partial S] \geq 1 - \sqrt{2\varepsilon}$. □

The above lemma has a corollary that if the sdp value is at least $(1 - \varepsilon)$ times the total weight of the edges (which is the value of the best possible cut), then the Goemans-Williamson algorithm also obtains a large fraction of this weight.

Corollary 1. Let $W := \sum_{e \in E} w_e$. If sdp $\geq (1 - \varepsilon)W$, then $\mathbf{Exp}[\mathsf{alg}_{\mathsf{GW}}] \geq W(1 - \sqrt{2\varepsilon})$.

Proof. This follows from the previous lemma and the concavity of the square root function. More precisely, let us define two random variables as follows: sample an edge (i,j) with probability $\frac{w(i,j)}{W}$

and let $X := \frac{1}{2}(1 - \mathbf{v}_i^\top \mathbf{v}_j)$ and $Y := \frac{\arccos(\mathbf{v}_i^\top \mathbf{v}_j)}{\pi}$. The premise of the algorithm implies $\mathbf{Exp}[X] = \frac{\mathsf{sdp}}{W} \geq 1 - \varepsilon$. The expected cut found by the Goemans-Williamson algorithm is $W \cdot \mathbf{Exp}[Y]$.

Now, Lemma 1 implies that $Y \geq 1 - \sqrt{2(1-X)}$ and so $\mathbf{Exp}[Y] \geq 1 - \mathbf{Exp}[\sqrt{2(1-X)}]$. Since $f(x) = \sqrt{2(1-x)}$ is a *concave* function, Jensen's inequality implies $\mathbf{Exp}[\sqrt{2(1-X)}] \leq \sqrt{\mathbf{Exp}[2(1-X)]} \leq \sqrt{2\varepsilon}$. And therefore, $\mathbf{Exp}[Y] \geq 1 - \sqrt{2\varepsilon}$, proving the lemma. $\square$

- ***A Different Interpretation.*** Here is a different interpretation of the SDP relaxation for the maximum cut. Imagine, a *prover* asserts that the optimum maxcut on a particular instance $\mathcal{I}$ is some number sdp, and in fact tells that not only is there on cut of this value, but rather a whole ***distribution*** $\mathcal{D}$ over such cuts. Indeed, considering cuts as Boolean vectors $\mathbf{Z} \in \{-1,+1\}^n$, the prover asserts that there is a distribution $\mathcal{D}$ over $\{-1,+1\}^n$ such that for any $\mathbf{Z} \sim \mathcal{D}$, the value of the cut is sdp. Of course, the prover is trying to cheat us; there may not be any such distribution, and our job is to get a "good" cut even if this prover is lying.

  If we, the *verifiers*, had full access to this distribution $\mathcal{D}$, then we would indeed also obtain cuts of the value sdp. There would be nothing left to do. Unfortunately, this prover isn't providing this distribution. Rather, they are willing to provide us with *moments* of this distribution. That is, for any $i \in V$, the prover can provide us $x_i = \mathbf{Exp}_{\mathbf{Z}\sim\mathcal{D}}[\mathbf{Z}_i]$; these are first moments, or the marginal expectations for each vertex i. Note that the prover could be lying about this as well, and we would like to put some "necessary conditions" on these first moments. One such constraint is that $-1 \leq x_i \leq 1$ since that is the range of $\mathbf{Z}_i$'s. And if one thinks about it, that is all we can say for the maximum cut problem. By the way, this is the interpretation that suffices for *any* linear programming relaxation for any of the problems we have seen in this course. All we get are "fractional" variables which can be interpreted as the first moments of individual variables. And then the linear constraints can be interpreted as expectations of linear constraints that the $\mathbf{Z}_i$'s must satisfy.

  Now suppose, that the prover is also willing us to provide *second moments* as well. That is, for any $i \in V, j \in V$, the prover is also willing to give us $x_{ij} = \mathbf{Exp}_{\mathbf{Z}\sim\mathcal{D}}[\mathbf{Z}_i\mathbf{Z}_j]$. Once again, how would we be convinced that these x_{ij}'s are indeed second moments? Is there a necessary condition on these x_{ij}'s? One trivial check is that x_{ii} better be 1 for all $i \in V$. If you remember the examples of PSD matrices we saw last time, one necessary condition on the $n \times n$ matrix $\mathbf{X}$ where $\mathbf{X}_{ij} = x_{ij}$, the purported second moments, is that $\mathbf{X} \succcurlyeq 0$. And if we club these two necessary conditions together, and note that the expected value of the cut $\mathbf{Z}$ drawn from $\mathcal{D}$ is precisely $\sum_{(u,v)\in E} w(u,v) \cdot \frac{1 - \mathbf{Exp}[\mathbf{Z}_i\mathbf{Z}_j]}{2}$, we get the SDP relaxation (Max Cut SDP).

- *Gaussian Rounding.* Does the interpretation above give any insight about rounding? Well, it really boils down to asking oneself: given access to the second moment matrix $\mathbf{X}$ from a purported distribution $\mathcal{D}$ over $\{-1,+1\}^n$, can we "learn" $\mathcal{D}$ itself? More precisely, can we sample from $\mathcal{D}$? If we could do so, then we would be solving our maxcut instances optimally. And so this task is also NP-hard. However, one can sample from a *different* true distribution $\mathcal{D}'$ with the same moment matrix $\mathbf{X}$ except that the support of $\mathcal{D}$' is not $\{-1,+1\}^n$ but rather $\mathbb{R}^n$. This is the *Gaussian Distribution* with moment matrix $\mathbf{X}$. We now describe this and how it really is just a reinterpretation of the Goemans-Williamson algorithm.

  First, recall that the one dimensional Gaussian, simply called a standard gaussian, $g \sim \mathcal{N}(0,1)$ with mean 0 and standard deviation 1 is distributed over $\mathbb{R}$ with probability distribution function $f(x) = e^{-x^2/2}$. Here is the reinterpreted Goemans-Williamson algorithm.

procedure GOEMANS-WILLIAMSON REINTERPRETED($X \succcurlyeq 0$):
    Obtain vectors $\mathbf{v}_i \in \mathbb{R}^d$ such that $\mathbf{X}_{ij} = \mathbf{v}_i^\top \mathbf{v}_j$.
    Let $V \in \mathbb{R}^{d \times n}$ be the matrix whose columns are $\mathbf{v}_i$'s.
    Sample independently $r_i \sim \mathcal{N}(0,1)$ for $1 \le i \le d$, and let $\mathbf{r} = (r_1, \ldots, r_d) \in \mathbb{R}^d$.
    Let $\mathbf{g} := V^\top \mathbf{r}$.
    $S = \{i : g_i \ge 0\}$.
    return $(S, V \setminus S)$.

Note that the $\mathbf{r}$ above is, after scaling by its length, precisely a point drawn uar from the unit sphere in d-dimensions. Also note that $g_i = \mathbf{v}_i^\top \mathbf{r}$, and therefore the set S is precisely the one returned by Goemans-Williamson algorithm. The analysis is exactly the same, and such a calculation was done a century ago in a different context.

To see this, first note that each g_i itself is a Gaussian random variable $\mathcal{N}(0,1)$. g_i is a linear combination of the r_i's, and a linear combination of standard gaussians is a standard gaussian with standard deviation equal to the $\sqrt{\|\mathbf{v}_i\|} = 1$. Secondly, note that these g_i's are *not* independent, but rather

$$\mathbf{Exp}[g_i g_j] = \mathbf{Exp}[(\mathbf{v}_i^\top \mathbf{r})(\mathbf{v}_j^\top \mathbf{r})] = \mathbf{v}_i^\top \mathbf{v}_j = \mathbf{X}_{ij}$$

The second inequality simply follows from the fact that r_i's are independent standard gaussians. Therefore, the distribution $\mathbf{g} = (g_1, \ldots, g_n)$ vector is precisely the (true) distribution $\mathcal{D}'$ whose moment matrix is $\mathbf{X}$. The only catch is that they are in $\mathbb{R}^n$ and not in $\{-1,+1\}^n$. The "rounding" algorithm is probably the first thing that comes to mind: use the "sign" of the g_i's to figure out which side the i's should fall. And this is precisely done above. Finally, the analysis of this algorithm depends on understanding how the signs of correlated random gaussians are distributed. And this was understood a long time back.

Lemma 2 (Sheppard's Formula)**.** Let G and H be two standard Gaussians which are ρ-correlated, that is, $\mathbf{Exp}[GH] = \rho$ for some $\rho \in [-1,+1]$. Then, $\mathbf{Pr}[\mathrm{sgn}(G) \neq \mathrm{sgn}(H)] = \frac{\arccos(\rho)}{\pi}$.

The proof of the above lemma is *precisely* the proof we sketched in the GW-analysis. One first writes (G,H) as a product of two independent standard gaussians (R_1,R_2) multiplied by a 2×2 correlation matrix. And then one notes that (R_1,R_2) is uniformly distributed over the circle (after scaling), and the probability that G and H are different signs is simply the probability that the random point on the circle splits the two points indicated by the rows of the correlation matrix. We leave these details out. It can also be done using calculus, but that is not the nicest way to do it.

Notes

The algorithm described here is from the seminal paper [3] by Goemans and Williamson, and the example of the 5-cycle can also be found in the paper. Soon after, this paper [2] by Feige and Schechtman gave an example which proved that $\alpha_{(\text{Max Cut SDP})} = \alpha_{\mathsf{GW}}$. Remarkably, a few years later, this paper [5] by Khot, Kindler, Mossel, and O'Donnell proved that under a certain yet unproven conjecture called the Unique Games Conjecture, it is NP-hard to obtain any approximation factor better than α_{GW}. This paper, and a follow-up paper [6] by Raghavendra opened up connections between SDPs and approximability. Without assuming UGC, it is known that Max Cut is NP-hard to approximate better than $\frac{16}{17} = 0.94$. This latter result is from another seminal paper [4] by Håstad.

The different interpretation is in fact a special case of a powerful way of looking at SDP relaxations. Indeed, by asking for higher moments from the prover, one can obtain stronger and stronger SDP relaxations. This way of looking at things is often called the “sum-of-squares” methodology, and is an active area of research obtaining exciting results. We point to the interested reader to this nice survey [1] by Barak and Steurer, and also to these lecture notes by Boaz Barak for a much deeper foray.

References

[1] B. Barak and D. Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of International Congress of Mathematicians (ICM)*, 2014.

[2] U. Feige and G. Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Structures Algorithms*, 20(3):403–440, 2002.

[3] M. X. Goemans and D. P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, pages 1115–1145, 1995.

[4] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[5] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing (SICOMP)*, 37(1):319–357, 2007.

[6] P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proc., ACM Symposium on Theory of Computing (STOC)*, pages 245–254, 2008.