6.S977 Lecture 5: Global Correlation Rounding

Last updated September 26, 2024

1 Constraint Satisfaction Problems

In this lecture, we will discuss another approach for rounding pseudoexpectations. In contrast to the Gaussian rounding strategy used in the Goemans-Williamson algorithm for max-cut, this time we will make use of higher-order moments of $\widetilde{\mathbb{E}}$. The setting we will consider will be that of 2-constraint satisfaction problems (2-CSPs), a generalization of max-cut.

Problem 1 (Constraint satisfaction problem). An instance of a 2-*CSP* is given by a collection of functions $\phi = (\phi_{ij})_{i,j\in[n]}$ where each $\phi_{ij} : [q] \times [q] \to \{0, 1\}$. The goal will be to maximize, over $x \in [q]^n$, the value of $\sum_{i,j} \phi_{ij}(x_i, x_j)$. We further denote the maximum by OPT.

It is not difficult to see that the above is a strict generalization of max-cut.

Example 2 (Max-cut as a CSP). Let G be a graph, and consider the 2-CSP with alphabet size $q = 2$, given by

$$\phi_{ij}(x_i, x_j) = \begin{cases} 0, & x_i = x_j, \\ 1, & x_i \neq x_j \end{cases}$$

for $ij \in E$, and $\phi_{ij} \equiv 0$ for $ij \notin E$. Then, for $x \in \{0, 1\}^n$, $\sum_{i,j} \phi_{ij}(x_i, x_j)$ is precisely the cut value associated to the cut $(\{i : x_i = 0\}, \{i : x_i = 1\})$.
More generally, for the alphabet being $[q]$ for $q \geq 2$, the above CSP attempts to find a q-coloring that violates the fewest edges, that is, minimizes the size of $\{ij \in E : x_i = x_j\}$.

The main result that we will discuss in this section, due to Barak-Raghavendra-Steurer [BRS11], is the following.

Theorem 3. Let ϕ be a 2-CSP on alphabet $[q]$. Then, for any $\varepsilon > 0$, there exists an algorithm that outputs x such that

$$\sum_{i,j} \phi_{ij}(x_i, x_j) \geq \mathsf{OPT} - \varepsilon n^2$$

that runs in time $(nq)^{O\left(\frac{\log q}{\varepsilon^2}\right)}$.

In fact, the running time in the above algorithm can be improved to $q^{O\left(\frac{\log q}{\varepsilon^2}\right)} \cdot n^{O(1)}$, which is polynomial in n. While we will not discuss the details thereof, this improvement involves designing a problem-specific semidefinite program solver, instead of blackbox using the ellipsoid algorithm that we saw in previous lectures. We refer the interested reader to [GS12] (also see [OGT13]).

The above algorithm provides a "polynomial time approximation scheme" (PTAS) for *dense* CSPs, where the optimum OPT is $\Omega(n^2)$. This is the case, for example, in max-cut on dense graphs with $\Omega(n^2)$ edges. While this was not the first approximation for max-cut on dense graphs, we will later see that the ideas here will generalize to certain non-dense graphs as well, such as expanders.

Given our discussion of max-cut and the sum-of-squares hierarchy, we now have a blueprint for designing optmization algorithms using sum-of-squares. We begin by encoding our constraints (here some proxy for $x_i \in [q]$) as low-degree polynomials, then search for pseudoexpectations that satisfy these constraints and maximize the objective polynomial.

How can we encode $x_i \in [q]$ for all $i \in [n]$? We shall consider a "one-hot" encoding of this constraint, by introducing variables y_{ia} for $i \in [n]$ and $a \in [q]$, indicating whether $x_i = a$, and enforce that the y_{ia} are in $\{0, 1\}$, and that exactly one of the $(y_{ia})_{a \in [q]}$ is equal to 1 (for any fixed $i \in [n]$). More concretely, consider the system of polynomials

$$\mathcal{P}_{[q]^n} := \left\{ \begin{array}{l} y_{ia}^2 = y_{ia} \text{ for all } i \in [n], a \in [q], \\ \sum_{a \in [q]} y_{ia} = 1 \text{ for all } i \in [n] \end{array} \right\}.$$

Furthermore, for $(y_i), (y_j)$ satisfying the above constraints, we abuse notation to denote by ϕ_{ij} the same function given by

$$\phi_{ij}(y_i, y_j) = \sum_{a,b \in [q]} \phi_{ij}(a, b) y_{ia} y_{jb}.$$

Note that this is a degree 2 polynomial in the (y_{ia}).

As usual, we try to maximize c such that there exists a pseudoexpectation $\widetilde{\mathbb{E}}$ such that

$$\widetilde{\mathbb{E}} \vDash \mathcal{P}_{[q]^n} \cup \left\{ \sum_{i,j \in [n]} \phi_{ij}(y_i, y_j) \geq c \right\}.$$

Obviously, the maximum such c is at least OPT.

An alternative similar expression would involve searching for a pseudoexpectation satisfying $\mathcal{P}_{[q]^n}$ that maximizes $\widetilde{\mathbb{E}} \sum_{i,j \in [n]} \phi_{ij}(y_i, y_j)$. However, this will cause issues later in the argument – in some sense, this maximizes the (pseudo)expected CSP value, while the expression we will use searches over (pseudo)distributions that are only supported on (y_{ia}) with large CSP value.

The rest of the game boils down to intelligently rounding the pseudoexpectation to a real distribution, and showing that this rounding does not lose too much in the objective value. In the Goemans-Williamson algorithm, we saw Gaussian rounding. In this lecture, we will see *global correlation rounding*.

2   Local Distributions and Independent Rounding

First, we will lay some groundwork, and in the process see another respect in which pseudoexpectations act like the moments of actual distributions on $\{0, 1\}^n$.

Lemma 4 (Local Distributions). Let $\widetilde{\mathbb{E}}$ be a pseudoexpectation of degree d. For every $|S| \subseteq [n]$ with $|S| \leq d/2$, there is a *real* distribution μ_S over $[q]^S$ such that for every $a_S = (a_i)_{i \in S}$ in $[q]^S$,

$$\Pr_{x_S | \mu_S} [x_i = a_i \text{ for all } i \in S] = \widetilde{\mathbb{E}} \prod_{i \in S} y_{ia_i}.$$

In other words, any pseudodistribution is "locally" a real distribution. We shall typically denote these "local distributions" by (μ_S).

Proof. We have for any $a_S = (a_i)_{i \in S}$ that

$$\widetilde{\mathbb{E}} \prod_{i \in S} y_{ia_i} = \widetilde{\mathbb{E}} \prod_{i \in S} y_{ia_i}^2 \geq 0.$$

Furthermore,

$$\sum_{a_S \in [q]^S} \widetilde{\mathbb{E}} \prod_{i \in S} y_{ia_i} = \widetilde{\mathbb{E}} \sum_{a_S \in [q]^S} \prod_{i \in S} y_{ia_i} = \widetilde{\mathbb{E}} \prod_{i \in S} \left(\sum_{a \in [q]} y_{ia} \right) = 1,$$

where the final inequality follows because $\widetilde{\mathbb{E}} \vDash \sum_{a \in [q]} y_{ia} = 1$ for all $i \in [n]$. This implies that the distribution μ_S defined as in the lemma statement is an actual distribution, as desired. ∎

Remark. This does *not* imply that there exists a single distribution μ over $[q]^n$ that is consistent with $\widetilde{\mathbb{E}}$. Indeed, were this the case, every pseudodistribution would be a real distribution. While there exist well-defined *local* distributions for every small set S, there is no guarantee that these local distributions can be "stitched together" to form a consistent local distribution.

As an exercise, show that these local distributions are consistent among themselves, in that for $T \subseteq S \subseteq [n]$ with $|S| \leq d/2$, the marginal of μ_S on T is equal to μ_T.

The above lemma suggests a simple rounding scheme: for each $i \in [n]$, independently draw $y_i \sim \mu_{\{i\}}$. This is not as silly as it may seem; in fact, many "randomized rounding" schemes for linear programs have this flavor and result in non-trivial algorithms.

Of course, the issue with this rounding scheme is that it loses all information about the higher-order correlations in $\widetilde{\mathbb{E}}$ – the joint distributions (y_i, y_j) will have independent coordinates, while the joint (2-local) distributions $\mu_{\{i,j\}}$ coming from $\widetilde{\mathbb{E}}$ need not be. Expressing the same concept in terms of pseudoexpectation values rather than local distributions, we expect

$$\Pr_{\substack{x_i \sim \mu_{\{i\}} \\ x_j \sim \mu_{\{j\}}}}[x_i = a, x_j = b] = \Pr_{x_i \sim \mu_{\{i\}}}[x_i = a] \cdot \Pr_{x_j \sim \mu_{\{j\}}}[x_j = b] = \widetilde{\mathbb{E}} y_{ia} \widetilde{\mathbb{E}} y_{jb}$$

and $\widetilde{\mathbb{E}} y_{ia} y_{jb}$ to be different.

Nevertheless, let us imagine for a moment that all such issues are absent, and proceed with an analysis. We have

$$\widetilde{\mathbb{E}} \sum_{i,j} \phi(x_i, x_j) = \sum_{i,j} \widetilde{\mathbb{E}} \phi(x_i, x_j)$$

$$= \sum_{i,j} \mathbb{E}_{(x_i, x_j) \sim \mu_{\{i,j\}}} \phi(x_i, x_j)$$

$$= \sum_{i,j} \mathbb{E}_{\substack{x_i \sim \mu_{\{i\}} \\ x_j \sim \mu_{\{j\}}}} \phi(x_i, x_j) - \sum_{i,j} \left(\mathbb{E}_{\substack{x_i \sim \mu_{\{i\}} \\ x_j \sim \mu_{\{j\}}}} \phi(x_i, x_j) - \mathbb{E}_{(x_i, x_j) \sim \mu_{\{i,j\}}} \phi(x_i, x_j) \right)$$

$$\geq \mathsf{OPT} - \sum_{i,j} d_{\mathrm{TV}} \left(\mu_{\{i\}} \otimes \mu_{\{j\}}, \mu_{\{i,j\}} \right).$$

Here, the final inequality follows from the property of the total variation distance that for distributions π, ν over Ω,

$$d_{\mathrm{TV}}(\pi, \nu) = \sup_{\substack{f: \Omega \to [0,1] \\ \text{measurable}}} \mathbb{E}_\pi[f] - \mathbb{E}_\nu[f],$$

and ϕ_{ij} takes values in $[0,1]$ ($\{0,1\}$, in fact).

As expected, the distribution not having the higher-order correlations of $\widetilde{\mathbb{E}}$ is causing issues. Indeed, the second term we are subtracting could be very large, potentially $\Omega(n^2)$.[1]

However, this presents us with a concrete goal! Can we somehow manipulate the pseudodistribution $\widetilde{\mathbb{E}}$ in a way that reduces the average correlation between pairs of coordinates, but preserves the objective value?

---

[1] Considering the distribution π over $\{0,1\}^2$ that takes $(0,1)$ with probability $1/2$ and $(1,0)$ with probability $1/2$, the total variation distance between π and $\pi_{\{1\}} \otimes \pi_{\{2\}}$ is $1/2$.

3 Conditioning

Now, we shall introduce another way in which pseudodistributions behave like real distributions – we can *condition* on events, provided that the events can be expressed by low-degree polynomials.

Lemma 5. Let $\widetilde{\mathbb{E}}$ be a degree d pseudodistribution over variables, such that $\widetilde{\mathbb{E}} y_{ia} > 0$. Then, the pseudodistribution $\widetilde{\mathbb{E}}[\cdot \mid y_{ia} = 1]$ defined by

$$\widetilde{\mathbb{E}}\left[p(y) \mid y_{ia} = 1\right] = \frac{\widetilde{\mathbb{E}} p(y) y_{ia}}{\widetilde{\mathbb{E}} y_{ia}}$$

is indeed a pseudodistribution of degree $d - 2$. Furthermore, this conditioned pseudodistribution satisfies any constraints that $\widetilde{\mathbb{E}}$ did.

We leave the proof of the above as an exercise to the reader.
The idea for the final rounding algorithm is as follows. Starting with some pseudoexpectation, we repeatedly pick a random coordinate in $[n]$, and pin the value of x_i according to the correct conditional (local) distribution (more concretely, we pick $a \sim \mu_{\{i\}}$, and condition on $y_{ia} = 1$). We will show that after some $\mathrm{poly}(\log q, 1/\varepsilon)$ steps, this algorithm outputs a pseudodistribution $\widetilde{\mathbb{E}}'$ (with local distributions (μ'_S)) with small pairwise correlations, in that

$$\sum_{i,j \in [n]} d_{\mathrm{TV}}\left(\mu'_{\{i,j\}}, \mu'_{\{i\}} \otimes \mu'_{\{j\}}\right) \leq \varepsilon n^2. \tag{1}$$

Why should this happen? Suppose instead that we are at a pseudodistribution that does not satisfy the above low-correlation guarantee. Then, for a randomly chosen coordinate i, we have $\sum_{j \in [n]} d_{\mathrm{TV}}\left(\mu'_{\{i,j\}}, \mu'_{\{i\}} \otimes \mu'_{\{j\}}\right) > \varepsilon n$ – the ith coordinate is non-trivially correlated with a lot of other coordinates j. Intuitively, this means that if you learn the value of the ith coordinate in a sample from μ, you would also learn a lot about the values of the other coordinates, thus driving down correlations.
To conclude the proof of Theorem 3, we shall prove the following lemma, known as the pinning lemma in statistical physics literature (where it was first introduced) and the global correlation lemma in sum-of-squares literature.

Lemma 6 (Global correlation lemma)**.** Let $\widetilde{\mathbb{E}}$ be an arbitrary pseudodistribution of degree at least $2T + 4$, for $T := O\left(\frac{\log q}{\varepsilon^2}\right)$. Let $\widetilde{\mathbb{E}}'$ be the (random) pseudodistribution obtained as follows. Starting with $\widetilde{\mathbb{E}}_0 = \widetilde{\mathbb{E}}$, for $0 \leq t \leq T$, do the following. If $\widetilde{\mathbb{E}}_t$ satisfies (1) or $t = T$, output $\widetilde{\mathbb{E}}' = \widetilde{\mathbb{E}}_t$. Otherwise, denoting by $\left(\mu_{\{i\}}^{(t)}\right)_{i \in [n]}$ the local distributions of $\widetilde{\mathbb{E}}_t$, pick a uniformly random $i \sim [n]$, $a_i \sim \mu_{\{i\}}$, and set $\widetilde{\mathbb{E}}_{t+1}[\cdot] = \widetilde{\mathbb{E}}_t[\cdot \mid y_{ia_i} = 1]$. Letting $(\mu')_{\substack{S \subseteq [n] \\ |S| \leq 2}}$ be the local distributions of $\widetilde{\mathbb{E}}'$, it holds that

$$\mathbb{E}_{\text{choices of pinning}} \sum_{i,j} d_{\mathrm{TV}}\left(\mu'_{\{i,j\}}, \mu'_{\{i\}} \otimes \mu'_{\{j\}}\right) \leq O(\varepsilon n^2).$$

4 Proof of the global correlation lemma

Before we start with the proof, we quickly cover the basics of information theory for the unfamiliar reader.

A Crash Course on Information Theory

Let Z, W be discrete random variables taking values in $[q]$, with probability density function $p_{Z,W} : [q] \times [q] \to \mathbb{R}_{\geq 0}$. Also let p_Z and p_W the density function of the marginal on Z and W. Then, the *entropy* of Z is defined by

$$H(Z) := \sum_{a \in [q]} -p_Z(a) \log p_Z(a),$$

where we use the convention $0 \log 0 = 0$, and all logs are in base e. The entropy of a random variable is a measure of "how random" it is, and is non-negative. Indeed, reinforcing this intuition that the entropy corresponds to randomness, it is maximized by the uniform distribution on $[q]$, which has an entropy of $\log q$. Given an event $\mathcal{E}$, we denote $H(Z \mid \mathcal{E}) = H(\{Z \mid \mathcal{E}\})$, where $\{Z \mid \mathcal{E}\}$ is the random variable obtaining by conditioning Z on $\mathcal{E}$. More generally, the *conditional entropy* is given by

$$H(Z \mid W) = \mathbb{E}_{w \sim W} H\left(Z \mid W = w\right).$$

This is a measure of how much randomness is "left in Z" after we condition on the value of W.
We also define the *mutual information*

$$I(Z; W) = H(Z) - H(Z \mid W),$$

which is a measure of how much knowing the value of W influences one's knowledge of the value of Z. The mutual information is non-negative and symmetric. Like the entropy, it is in $[0, \log q]$.
Finally, a consequence of Pinsker's inequality is that

$$d_{\mathrm{TV}}\left(P_{Z,W}, P_Z \otimes P_W\right) \leq \sqrt{\frac{1}{2} I(Z; W)}. \tag{2}$$

Let us prove Lemma 6. For $0 \leq t \leq T$, let $\widetilde{\mathbb{E}}_t$ be the random pseudodistribution obtained after t rounds of pinning. We now introduce some notation, some of which we will abuse. Let $X_i^{(t)}$ denote the local distribution of $\widetilde{\mathbb{E}}_t$ on x_i; we shall also use $(X_i^{(t)}, X_j^{(t)})$ to mean the local distribution on (x_i, x_j) according to $\widetilde{\mathbb{E}}_t$. This abuse is not inconsistent (in that it is okay to use the same letter $X_i^{(t)}$ in both places) due to the exercise in the remark after Lemma 4. In particular, since these are legitimately random variables, we may talk about the entropy of $H(X_i^{(t)})$ and the mutual information $I(X_i^{(t)}; X_j^{(t)})$.

Consider the potential function

$$\psi^{(t)} := \mathbb{E}_{\substack{\text{choices of pinnings} \\ \text{up to time } t}} \left[\mathbb{E}_{i \sim [n]} H(X_i^{(t)})\right].$$

We will show that this potential function is decaying, and if something looking like the average pairwise correlations is large, it decays fast. Towards this, define the *global information*

$$\mathrm{GI}^{(t)} = \mathbb{E}_{\substack{\text{choices of pinnings} \\ \text{up to time } t}} \mathbb{E}_{i,j \sim [n]} I(X_i^{(t)}; X_j^{(t)}).$$

We will show that

$$\psi^{(t)} - \psi^{(t+1)} = \mathrm{GI}^{(t)}.$$

Indeed,

$$\mathrm{GI}^{(t)} = \mathbb{E}_{\substack{\text{choices of pinnings} \\ \text{up to time } t}} \mathbb{E}_{i,j \sim [n]} H(X_i^{(t)}) - H(X_i^{(t)} \mid X_j^{(t)})$$

$$= \mathbb{E}_{\substack{\text{choices of pinnings} \\ \text{up to time } t}} \mathbb{E}_{i \sim [n]} H(X_i^{(t)}) - \mathbb{E}_{\substack{\text{choices of pinnings} \\ \text{up to time } t}} \mathbb{E}_{\substack{j \sim [n] \\ a_j \sim \mu_{\{j\}}^{(t)}}} \mathbb{E}_{i \sim [n]} H(X_i^{(t)} \mid X_j^{(t)} = a_j).$$

The first quantity is equal to $\psi^{(t)}$ by definition. Furthermore, note that the second expectation in the second quantity exactly describes the pinning in the $(t+1)$th step! Therefore,

$$\mathrm{GI}^{(t)} = \underset{\substack{\text{choices of pinnings} \\ \text{up to time } t}}{\mathbb{E}} \mathbb{E}_{i\sim[n]} H(X_i^{(t)}) - \underset{\substack{\text{choices of pinnings} \\ \text{up to time } t+1}}{\mathbb{E}} \mathbb{E}_{i\sim[n]} H(X_i^{(t+1)}) = \psi^{(t)} - \psi^{(t+1)}.$$

Now, due to the non-negativity of entropy, $\psi^{(T)} \geq 0$, and due to the boundedness of entropy, $\psi^{(0)} \leq \log q$. It follows that

$$\sum_{0 \leq t \leq T} \mathrm{GI}^{(t)} \leq \log q.$$

Therefore,

$$\frac{\log q}{T} \geq \mathbb{E}_{\text{choices of pinnings}} \frac{1}{T} \sum_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} I(X_i^{(t)}; X_j^{(t)})$$

$$\geq \mathbb{E}_{\text{choices of pinnings}} \min_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} I(X_i^{(t)}; X_j^{(t)})$$

$$\overset{(2)}{\geq} 2 \cdot \mathbb{E}_{\text{choices of pinnings}} \min_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} d_{\mathrm{TV}}\left(\mu_{\{i,j\}}^{(t)}, \mu_{\{i\}}^{(t)} \otimes \mu_{\{j\}}^{(t)}\right)^2.$$

$$\geq 2 \cdot \left(\mathbb{E}_{\text{choices of pinnings}} \min_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} d_{\mathrm{TV}}\left(\mu_{\{i,j\}}^{(t)}, \mu_{\{i\}}^{(t)} \otimes \mu_{\{j\}}^{(t)}\right)\right)^2,$$

where the final inequality is Jensen's inequality applied twice. Recalling that $T = \Omega\left(\frac{\log q}{\varepsilon^2}\right)$,

$$\mathbb{E}_{\text{choices of pinning}} \min_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} d_{\mathrm{TV}}\left(\mu_{\{i,j\}}^{(t)}, \mu_{\{i\}}^{(t)} \otimes \mu_{\{j\}}^{(t)}\right) \leq O(\varepsilon),$$

so for the output distribution $\widetilde{\mathbb{E}}'$ with local distributions (μ_S'),

$$\mathbb{E}_{\text{choices of pinning}} \mathbb{E}_{i,j\sim[n]} d_{\mathrm{TV}}\left(\mu_{\{i,j\}}', \mu_{\{i\}}' \otimes \mu_{\{j\}}'\right) \leq \mathbb{E}_{\text{choices of pinning}} \max\left\{\varepsilon, \min_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} d_{\mathrm{TV}}\left(\mu_{\{i,j\}}^{(t)}, \mu_{\{i\}}^{(t)} \otimes \mu_{\{j\}}^{(t)}\right)\right\}$$

$$\leq \varepsilon + \mathbb{E}_{\text{choices of pinning}} \min_{0 \leq t \leq T} \mathbb{E}_{i,j\sim[n]} d_{\mathrm{TV}}\left(\mu_{\{i,j\}}^{(t)}, \mu_{\{i\}}^{(t)} \otimes \mu_{\{j\}}^{(t)}\right)$$

$$\leq O(\varepsilon).$$

This completes the proof of Lemma 6, and thus Theorem 3. ∎

An interesting aspect to the above analysis is that it does *not* show that $\mathrm{GI}^{(t)}$ is monotone. This is the reason for the strange form of Lemma 6 – aesthetically, a nicer algorithm would involve just outputting $\widetilde{\mathbb{E}}_T$.

5   Beyond dense 2-CSPs

The above algorithmic idea, while explained in the context of dense 2-CSPs, goes far beyond. For starters, a similar analysis works for dense k-CSPs, which involves functions $\phi_{i_1,\ldots,i_k} : [q]^k \to \{0,1\}$. Here, there exists an algorithm that outputs an assignment x with

$$\sum \phi_{i_1,\ldots,i_k}(x_1,\ldots,x_k) \geq \mathsf{OPT} - \varepsilon n^k$$

that runs in time $n^{\mathrm{poly}(\log q, 1/\varepsilon, k)}$.

In fact, the ideas here also work for some important classes of non-dense 2-CSPs. An important family of such CSPs is given by expanders. Recall that a graph G with adjacency matrix A_G and diagonal degree matrix D is said to be a *two-sided spectral expander* if the normalized adjacency matrix $D^{-1/2} A_G D^{-1/2}$, which has eigenvalues $(\lambda_i)_{i \in [n]}$ in $[-1, 1]$ and top eigenvalue equal to 1, is such that $|\lambda_i| \leq \varepsilon^{O(1)}$ for $2 \leq i \leq n$.

Consider a CSP supported on such a graph, in that $\phi_{ij} \equiv 0$ if ij is not an edge. For such graphs, it turns out the algorithm discussed here outputs an assignment x such that

$$\sum_{ij \in G} \phi_{ij}(x_i, x_j) \geq \mathsf{OPT} - \varepsilon |E(G)|.$$

The situation is even more general. Suppose that G is a graph whose normalized adjacency matrix, with eigenvalues $(\lambda_i)_{i \in [n]}$, is such that there are at most r eigenvalues with $|\lambda_i| \geq \varepsilon^{O(1)}$. Then, a similar algorithm outputs x such that

$$\sum_{ij \in G} \phi_{ij}(x_i, x_j) \geq \mathsf{OPT} - \varepsilon |E(G)|,$$

running in time $n^{\mathrm{poly}(\log q, 1/\varepsilon, r)}$. This quantity r is sometimes called the "threshold rank" of G, and much effort has been spent trying to show that low-threshold rank graphs behave like dense graphs in various senses. While the details of this analysis are largely similar those we have presented, we omit them as their explanation would require a detour into spectral graph theory – we refer the interested reader to

References

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 472–481. IEEE, 2011.

[GS12]   Venkatesan Guruswami and Ali Kemal Sinop. Faster sdp hierarchy solvers for local rounding algorithms. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 197–206. IEEE, 2012.

[OGT13] Shayan Oveis Gharan and Luca Trevisan. A new regularity lemma and faster approximation algorithms for low threshold rank graphs. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 303–316. Springer, 2013.