Global Correlation Rounding

Alert: these notes are a work in progress, and have not been subjected to the usual scrutiny reserved for formal publications!

We will describe another approach for rounding pseudoexpectations. By contrast to the Gaussian rounding strategy used for max-cut, this time we will make use of higher-order moments of $\tilde{\mathbb{E}}$. We will continue to use max-cut as our running example, but the ideas we will see are useful well beyond the setting of max-cut.

Recall that for an n-vertex graph G, we have defined $G(x) = \sum_{i \sim j}(x_i - x_j)^2$ to be the function which counts the number of edges of G cut by some $0/1$ vector x. Our goal will be to prove the following theorem:

Theorem 1 (Barak-Raghavendra-Steurer): Let G be an n-vertex graph containing $\Omega(n^2)$ edges. Then for $d \gg 1/\varepsilon^{O(1)}$, $\vdash_d G(x) \le (1 + \varepsilon) \max_y G(y)$.

Here we have introduced a new notational shorthand: $\vdash_d f \le g$ is shorthand for $\vdash_d g - f \ge 0$.

Let us compare Theorem 1 to the theorem we previously proved for degree-2 SoS upper bounds on G: we have traded the constant $1/0.878$ for $1 + \varepsilon$, at the expense of using SoS proofs of somewhat higher degree, $\mathrm{poly}(1/\varepsilon)$, and a fairly strong assumption on the graph G, that it is dense. Later, we will see how to weaken the second assumption. Once again, implicit in the proof will be an algorithm for actually finding cut which is at most $(1 + \varepsilon)$ times the size of the maximum cut, rather than just approximating the max-cut value.

Pseudoexpectations, Local Distributions, and Conditioning

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

Lemma 2 (Local Distributions): Let $\tilde{\mathbb{E}}$ be a pseudoexpectation of degree d. For every $|S| \subseteq [n]$ with $|S| \le d/2$, there is a distribution $\mu_S : \{0, 1\}^k \to \mathbb{R}$ such that for every $T \subseteq S$, $\tilde{\mathbb{E}}x^T = \mathbb{E}_{y \sim \mu_S} y^T$.

Proof: Fixing S, consider $\tilde{\mathbb{E}}$ restricted to the variables x_i for $i \in S$; there are at most $d/2$ of them and $\tilde{\mathbb{E}}$ has degree d. Therefore, $\tilde{\mathbb{E}}$ represents the moments of an actual distribution μ_S on $\{0, 1\}^{|S|}$. QED.

Another respect in which pseudoexpectations act like (moments of) distributions is that they can be *conditioned* on events (as long as those events are sufficiently simple, by which we mean that they can be described by a low-degree polynomial).

Concretely, let $\tilde{\mathbb{E}}$ be a pseudoexpectation of degree d, and let $i \in [n]$ be such that $\tilde{\mathbb{E}}x_i > 0$. If $\tilde{\mathbb{E}}$ were the moments of an actual distribution, x_i would be a variable which has nonzero probability of being equal to 1, and we could condition on that event. To "pseudo-condition" on the event, we define a new pseudoexpectation:

1

$$\tilde{\mathbb{E}}[\cdot \,|\, x_i = 1] \ : \ \mathbb{R}[x]_{\leq d-2} \to \mathbb{R}$$

via

$$\tilde{\mathbb{E}}[x^S \,|\, x_i = 1] := \frac{\tilde{\mathbb{E}}[x^S x_i]}{\tilde{\mathbb{E}}[x_i]}$$

Similarly, if $\tilde{\mathbb{E}}[(1 - x_i)] > 0$, we can define

$$\tilde{\mathbb{E}}[x^S \,|\, x_i = 0] := \frac{\tilde{\mathbb{E}}[x^S(1 - x_i)]}{\tilde{\mathbb{E}}[(1 - x_i)]}$$

Lemma 3 (Conditioning): $\tilde{\mathbb{E}}[\cdot \,|\, x_i = 1]$ and $\tilde{\mathbb{E}}[\cdot \,|\, x_i = 0]$ are pseudoexpectations of degree $d - 2$.

Proof: Exercise (just check the definition of a pseudoexpectation).

For some intuition, let us sanity check that $\tilde{\mathbb{E}}[\cdot \,|\, x_i = 1]$ "acts like" moments of a distribution where x_i always assumes the value 1. For instance, it should be the case that for any polynomial p, $\tilde{\mathbb{E}}[p(x)x_i \,|\, x_i = 1] = \tilde{\mathbb{E}}[p(x) \,|\, x_i = 1]$. Luckily, it is easy to check that this is the case.

Independent Rounding

To prove Theorem 1, we will describe an algorithm which takes a pseudoexpectation $\tilde{\mathbb{E}}$ of degree d and finds $y \in \{0, 1\}^n$ such that $G(y) \geq (1 - \varepsilon)\tilde{\mathbb{E}}G(x)$. As before, we will need a "rounding strategy" to do this.

Here's an even simpler idea than the Gaussian rounding approach we tried before. The pseudoexpectation $\tilde{\mathbb{E}}$ specifies 1-wise marginal distributions for each coordinate, where coordinate i is 1 with probability $\tilde{\mathbb{E}}x_i$ and otherwise 0. (Exercise: sanity check that $\tilde{\mathbb{E}}x_i \in [0, 1]$ so that this is well defined.) We could simply sample each coordinate independently from according to these distributions. (This is not so dumb as it seems: in fact, many "randomized rounding" schemes for linear programs have this flavor and lead to nontrivial algorithms.)

Of course, the resulting random vector y will not necessarily have the same higher-order correlations as $\tilde{\mathbb{E}}$ – that is, the joint distributions (y_i, y_j) will have independent coordinates, while the joint (2-local) distributions μ_{ij} coming from $\tilde{\mathbb{E}}$ need not be. (Expressing the same concept in terms of pseudoexpectation values rather than local distributions: we expect $\mathbb{E}y_i y_j = \mathbb{E}y_i \mathbb{E}y_j$ and $\tilde{\mathbb{E}}x_i x_j$ to be different.)

But, for a thought experiment, let us imagine that we have gotten lucky, and the local distributions μ_{ij} are close to independent. Concretely, suppose

2

$$\sum_{i,j} |\mu_{ij} - \mu_i \otimes \mu_j|_{TV} \le \delta n^2.$$

Here, $|\cdot|_{TV}$ denotes total variation distance, and $\mu_i \otimes \mu_j$ is the product of the two 1-local distributions extracted from the pseudoexpectation $\tilde{\mathbb{E}}$. Let's look at what happens when we use the independent rounding strategy on $\tilde{\mathbb{E}}$.

$$
\begin{aligned}
\mathbb{E}_y G(y) &= \sum_{i \sim j} \Pr(y_i \ne y_j) \\
&= \sum_{i \sim j} \Pr_{x \sim \mu_i \otimes \mu_j} (x_i \ne x_j) \\
&\ge \sum_{i \sim j} \Pr_{x \sim \mu_{ij}} (x_i \ne x_j) - |\mu_{ij} - \mu_i \otimes \mu_j|_{TV} \\
&\ge \sum_{i \sim j} \Pr_{x \sim \mu_{ij}} (x_i \ne x_j) - \delta n^2 \\
&= \tilde{\mathbb{E}} G(x) - \delta n^2.
\end{aligned}
$$

In this case (still a thought experiment), independent rounding has gone well, incurring just an additive δn^2 loss! But how can we arrange for $\tilde{\mathbb{E}}$ to satisfy the approximate 2-wise independence condition?

Reducing Global Correlation by Conditioning

The following lemma captures the key idea we'll use to arrange for $\tilde{\mathbb{E}}$ to satisfy the condition needed for independent rounding. It was discovered independently by Montanari, in the context of statistical physics and later by Barak, Raghavendra, Steurer, and Tan, in the SoS context. The name "pinning lemma" comes from the statistical physics literature, where the act of conditioning on an event $x_i = 1$ is referred to as "pinning" the value of x_i to 1.

Lemma 4 (Pinning Lemma): Let $\tilde{\mathbb{E}}$ be a degree-d pseudoexpectation on n variables, with $d \ll n$. There exists $t \le d - 2$ such that if $S \subseteq [n]$ is a random set of coordinates with $|S| = t$ and y_S is a sample from the local distribution μ_T induced on $\{0,1\}^t$ by $\tilde{\mathbb{E}}$,

$$\mathbb{E}_{S, y_S} \sum_{i,j} |\mu_{ij \,|\, y_S} - \mu_{i \,|\, y_S} \otimes \mu_{j \,|\, y_S}|_{TV} \le O(n^2/\sqrt{d}),$$

where $\mu_{T \,|\, y_S}$ denotes the local distribution on T of the conditional pseudoexpectation $\tilde{\mathbb{E}}[\cdot \,|\, x_S = y_S]$ (which we define by conditioning iteratively on the coordinates).

Proof idea: Before we prove the lemma, let's explain the main idea, as usual drawing on intuition we cultivate by thinking of $\tilde{\mathbb{E}}$ as representing the moments of an actual distribution. Suppose we have such a distribution μ, and that the coordinates of μ are quite correlated, $\sum_{i,j} |\mu_{ij} - \mu_i \otimes \mu_j|_{TV} \gg \delta n^2$. We can rewrite this as

$$\mathbb{E}_{i \sim [n]} \sum_{j \leq n} |\mu_{ij} - \mu_i \otimes \mu_j|_{TV} \gg \delta n,$$

so for a randomly chosen coordinate i, we expect to have $\sum_{j \leq n} |\mu_{ij} - \mu_i \otimes \mu_j|_{TV} \gg \delta n$ – the i-th coordinate is nontrivially correlated with lots of other coordinates j. Intuitively, if you learned the value of the i-th coordinate in a sample from μ, you would therefore also learn a lot about the values of the other coordinates!

A little more formally, conditioning on the value of coordinate i should cause the *entropy* of the other coordinates to decrease. There is only so much entropy available, meaning that if we repeat this conditioning procedure, eventually we should have $\sum_{i,j} |\mu'_{ij} - \mu'_i \otimes \mu'_j|_{TV} \leq \delta n^2$, where μ' is some conditioning of μ on the values of some coordinates. It turns out that this will happen after at most $O(1/\delta^2)$ conditioning steps.

The cleanest way to express these ideas formally goes via entropy and mutual information. There are many excellent resources to learn elementary information theory, so we'll just state the basic facts we need to use here.

Let $H(X)$ denote the entropy of a (discrete) random variable X and $I(X;Y)$ the mutual information between X and Y.

Lemma (Entropy Facts):

(1) If X, Y are jointly-distributed $\{0, 1\}$-valued random variables, $H(X), H(Y), I(X;Y) \in [0, 1]$.
(2) (Pinsker's inequality) For jointly distributed random variables X, Y, denote by μ the joint distribution and μ_X, μ_Y the marginal distributions of X and Y. Then $|\mu - \mu_X \otimes \mu_Y|_{TV} \leq \sqrt{\frac{1}{2} I(X;Y)}$.

Proof of Lemma 4: For $s \leq d - 2$, let us define a (random) sequence of pseudo-expectations $\tilde{\mathbb{E}} = \tilde{\mathbb{E}}^{(0)}, \tilde{\mathbb{E}}^{(1)}, \ldots, \tilde{\mathbb{E}}^{(d-2)}$, where $\tilde{\mathbb{E}}^{(s)}$ is obtained from $\tilde{\mathbb{E}}^{(s-1)}$ by choosing a random $i \in [n]$ which has not been chosen in a previous step, sampling y_i from the local distribution $\mu_i^{(s-1)}$, and setting $\tilde{\mathbb{E}}^{(s)} = \tilde{\mathbb{E}}^{(s-1)}[\cdot \,|\, x_i = y_i]$.

(Exercise: check that the following alternative definition gives the same distribution for $\tilde{\mathbb{E}}^{(s)}$. Choose $S \subseteq [n]$ with $|S| = s$ at random and sample y_S from the local distribution μ_S induced by $\tilde{\mathbb{E}}$. Then let $\tilde{\mathbb{E}}^{(s)} = \tilde{\mathbb{E}}[\cdot \,|\, x_S = y_S]$.)

Consider the *global information* for the s-th pseudoexpectation:

$$\text{global}_s := \mathbb{E}_{i,j \in [n]} I(X_i^{(s)}; X_j^{(s)})$$

where, in the i,j-th term of the sum, $X_i^{(s)}, X_j^{(s)}$ denote a jointly distributed sample from the 2-local distribution $\mu_{ij}^{(s)}$ induced by $\tilde{\mathbb{E}}^{(s)}$.

If, for some s, we have $\mathbb{E}\text{global}_s \leq \delta$, then by convexity of $\sqrt{\cdot}$ and Pinsker's inequality, we have $\mathbb{E}\sum_{i,j}|\mu_{ij}^{(s)} - \mu_i^{(s)} \otimes \mu_j^{(s)}|_{TV} \leq \sqrt{\delta}n^2$. For $\delta = O(1/d)$, this would complete the proof, so we just need to show that $\mathbb{E}\text{global}_s \leq O(1/d)$ for some $s \leq d - 2$.

We introduce the following potential function tracking the amount of entropy in 1-local distributions:
$$\phi^{(s)} := \mathbb{E}\mathbb{E}_{i \in [n]}H(X_i^{(s)}).$$

(The outer expectation averages over all the randomness in the choice of s and the values to be conditioned on.) The key claim, proved below, is that for $d \ll n$,:

$$\phi^{(s)} - \phi^{(s+1)} \leq \Omega(\mathbb{E}\text{global}_s).$$

We know $0 \leq \phi^{(s)} \leq 1$, using $H(X) \in [0,1]$ for a binary random variable. So,

$$1 \geq \phi^{(0)} - \phi^{(d-2)} \geq \Omega\left(\sum_{s \leq d-2} \mathbb{E}\text{global}_s\right).$$

Finally, since mutual information is nonnegative, $\text{global}_s \geq 0$ for all s. So, some term in the sum is at most $O(1/d)$. QED

Proof of claim: By definition, $I(X_i^{(s-1)}; X_j^{(s-1)}) = H(X_i^{(s-1)}) - H(X_i^{(s-1)} \mid X_j^{(s-1)})$. In the s-th step, we choose some index j_s to condition on, having already conditioned on indices $j_1, \ldots, j_{s-1}$. So,

$$\phi^{(s-1)} - \phi^{(s)} = \mathbb{E}_{j_1,\ldots,j_{s-1}}\mathbb{E}_{j \sim [n]\backslash j_1,\ldots,j_{s-1}}\mathbb{E}_{i \sim [n]}H(X_i^{(s-1)} - H(X_i^{(s-1)}|X_j^{(s-1)})$$
$$= \mathbb{E}_{j_1,\ldots,j_{s-1}}\mathbb{E}_{i \sim [n]\backslash j_1,\ldots,j_{s-1}}I(X_i^{(s-1)}; X_j^{(s-1)})$$
$$\leq \frac{n}{n-(s-1)}\mathbb{E}\text{global}_{s-1}.$$

QED.

Putting it together

Now we're ready to prove Theorem 1.

Proof of Theorem 1 Suppose $\tilde{\mathbb{E}}$ is a degree-d pseudoexpectation. Following the procedure in Lemma 4, we know that we can obtain a (random) conditional pseudoexpectation $\tilde{\mathbb{E}}'$, with local distributions μ' satisfying

$$\mathbb{E}\sum_{i,j \leq n}|\mu'_{ij} - \mu'_i \otimes \mu'_j|_{TV} \leq O(n^2/\sqrt{d}).$$

5

Furthermore, $\mathbb{E}\tilde{\mathbb{E}}'G(x) = \tilde{\mathbb{E}}G(x)$, so if we view $\tilde{\mathbb{E}}'G(x)$ itself as a random variable bounded between 0 and n^2, we have $\Pr(\tilde{\mathbb{E}}'G(x) < (1-\alpha)\tilde{\mathbb{E}}G(x)) \le \alpha\tilde{\mathbb{E}}G(x)/n^2$, for any $\alpha > 0$. By Markov's inequality, we can assume the local distributions μ' satisfy

$$\sum_{i,j \le n} |\mu'_{ij} - \mu'_i \otimes \mu'_j|_{TV} \le \frac{1}{\alpha} \cdot O(n^2/\sqrt{d})$$

with probability $1 - \alpha$. Choosing $\alpha = 1/d^{1/4}$, there is $\tilde{\mathbb{E}}'$ with local distributions μ' such that

$$\sum_{i,j \le n} |\mu'_{ij} - \mu'_i \otimes \mu'_j|_{TV} \le O(n^2/d^{1/4})$$

and $\tilde{\mathbb{E}}'G(x) \ge (1 - d^{-1/4})\tilde{\mathbb{E}}G(x)$. Putting this together with our previous analysis of independent rounding, we can see that independent rounding produces a cut y with $\mathbb{E}G(y) \ge \tilde{\mathbb{E}}G(x) - O(n^2/d^{1/4})$. QED.

Local-to-Global: Extending to Expanders

It's possible to take these ideas well beyond Theorem 1 – generalizing to constraint satisfaction problems beyond max-cut, as well as past the case of dense graphs. We'll show one important generalization here, to *expander graphs*.

For this section, we consider Δ-regular graphs on n vertices. Let A_G be the normalized adjacency matrix of such a graph G, so $\lambda_1(A) = 1$.

Theorem 2: For every Δ-regular graph G and every even d, $\vdash_d G(x) \le (1 + d^{-\Omega(1)} + \lambda_2^{\Omega(1)}) \max_y G(y)$.

In the case of dense graphs, we could show that independent rounding succeeded when the quantity

$$(1) \qquad \mathbb{E}_{i,j \sim [n]} |\mu_{ij} - \mu_i \otimes \mu_j|_{TV}$$

was small. In sparse graphs, this is no longer the case. But, by inspecting the foregoing analysis, we can see that independent rounding will work fine if we can bound

$$(2) \qquad \mathbb{E}_{i \sim j} |\mu_{ij} - \mu_i \otimes \mu_j|_{TV},$$

that is, measuring how close the 2-local distributions are to being product when averaged only over edges in G.

For certain graphs – expanders among them – it is possible to show that the pinning operation actually leads to a pseudoexpectation for which (2) is small, via a "local to global" lemma. We'll need a definition first – here it is convenient to switch to the $\{\pm 1\}$ basis for the hypercube, which is done implicitly in the definition statement.

Definition (Global and Local Correlation) Let $\tilde{\mathbb{E}}$ be a degree-2 pseudoexpectation, and for each $i \in [n]$ let $y_i = y_i(x) = 2x_i - 1$. Let $Cov_{ij} = \tilde{\mathbb{E}}y_iy_j - \tilde{\mathbb{E}}y_i\tilde{\mathbb{E}}y_j$. We define the *global correlation* as

$$\text{global-corr} = \mathbb{E}_{i,j \sim [n]} Cov_{ij}^2$$

and, for a graph G, the *local correlation* as

$$\text{local-corr} = \mathbb{E}_{i \sim j} Cov_{ij}^2.$$

The following lemma relates covariance to information (and is the reason we switched from the x's to the y's).

Lemma (Covariance versus Information): For $\{\pm 1\}$-valued random variables X, Y, $Cov(X,Y)^2 \leq O(I(X;Y))$.

Proof: See Raghavendra-Tan, Fact B.5.

Lemma (Local-to-Global): Let $\tilde{\mathbb{E}}$ be a degree-4 pseudoexpectation, G a Δ-regular graph with normalized adjacency matrix A_G whose eigenvalues are $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Then local-corr $\leq$ global-corr $+ O(\lambda_2)$.

In the process of proving the local-to-global lemma, we'll see a new kind of SoS proof, built out of the eigenvalues/eigenvectors of a PSD matrix. The following argument will use some basic spectral graph theory – Dan Spielman's book is an excellent reference for this material.

Proof of local-to-global lemma: Let us recall a few basic facts about the quadratic form $v^\top A_G v$. Since A_G is regular, its top eigenvector is the all-1's, so we can write

$$\frac{1}{\Delta} \sum_{i \sim j} v_i v_j = v^\top A_G v$$

$$= \frac{1}{n} \sum_{i,j} v_i v_j + \sum_{1 < j \leq n} \lambda_j(A_G) \langle v, w_j \rangle^2$$

where w_j is (unit-norm) the eigenvector of A_G associated to eigenvalue $\lambda_j(A_G)$.

Now let's write local correlation as

$$\text{local-corr} = \mathbb{E}_{i \sim j} (\tilde{\mathbb{E}} y_i y_j - \tilde{\mathbb{E}} y_i \tilde{\mathbb{E}} y_j)^2$$

$$= \frac{1}{n} \cdot \frac{1}{\Delta} \sum_{i \sim j} (\tilde{\mathbb{E}} y_i y_j - \tilde{\mathbb{E}} y_i \tilde{\mathbb{E}} y_j)^2$$

$$= \frac{1}{n} \cdot \frac{1}{\Delta} \tilde{\mathbb{E}} \left[\sum_{i \sim j} (y_i - \tilde{\mathbb{E}} y_i)(y_j - \tilde{\mathbb{E}} y_j)(y_i' - \tilde{\mathbb{E}} y_i')(y_j' - \tilde{\mathbb{E}} y_j) \right]$$

where y_i' and y_j' are independent copies of y_i and y_j. (Exercise: check that there is a well-defined pseudoexpectation $\tilde{\mathbb{E}}$ on variables $y_1, \ldots, y_n, y_1', \ldots, y_n'$.)

The last expression we can rewrite again as

$$\frac{1}{n} \cdot \frac{1}{\Delta} \tilde{\mathbb{E}} \left[\sum_{i \sim j} (y_i - \tilde{\mathbb{E}} y_i)(y_j - \tilde{\mathbb{E}} y_j)(y_i' - \tilde{\mathbb{E}} y_i')(y_j' - \tilde{\mathbb{E}} y_j) \right]$$

$$= \frac{1}{n^2} \tilde{\mathbb{E}} \sum_{i,j \in [n]} (y_i - \tilde{\mathbb{E}} y_i)(y_j - \tilde{\mathbb{E}} y_j)(y_i' - \tilde{\mathbb{E}} y_i')(y_j' - \tilde{\mathbb{E}} y_j) + \tilde{\mathbb{E}} \frac{1}{n} \sum_{1 < j \leq n} \lambda_j(A_G) \langle v(y,y'), w_j \rangle^2$$

$$= \text{global-corr} + \tilde{\mathbb{E}} \frac{1}{n} \sum_{1 < j \leq n} \lambda_j(A_G) \langle v(y,y'), w_j \rangle^2$$

where $v(y,y')$ is the vector-valued polynomial given by $v(y,y')_i = (y_i - \tilde{\mathbb{E}} y_i)(y_i' - \tilde{\mathbb{E}} y_i')$.

Finally, we'll use the following claim, established below:

$$\vdash_4 \frac{1}{n} \sum_{1 < j \leq n} \lambda_j(A_G) \langle v(y,y'), w_j \rangle^2 \leq O(\lambda_2).$$

All together, we've obtained local-corr $\leq$ global-corr $+ O(\lambda_2)$.

Proof sketch of claim: Use PSDness of $\lambda_2 I - \sum_{1 < j \leq n} \lambda_j(A_G) w_j w_j^\top$ together with $\vdash_4 \|v(y,y')\|^2 \leq n$.

Now let's put things together to prove Theorem 2.

Proof sketch of Theorem 2: Let $\tilde{\mathbb{E}}$ be a degree-d pseudoexpectation. Following the proof of Theorem 1', we can assume there is a conditioning $\tilde{\mathbb{E}}'$ of $\tilde{\mathbb{E}}$, with local distributions μ', such that:

- $\mathbb{E}_{i,j \in [n]} I(X_i', X_j') \leq d^{-\Omega(1)}$
- $\tilde{\mathbb{E}}' G(x) \geq (1 - d^{-\Omega(1)}) \tilde{\mathbb{E}} G(x)$.

The first point implies that $\mathbb{E}_{i,j \in [n]} Cov_{ij}^2 \leq d^{-\Omega(1)}$, so by the local-to-global lemma, $\mathbb{E}_{i \sim j} Cov_{ij}^2 \leq d^{-\Omega(1)} + O(\lambda_2)$.

Exercise: show that $|\mu_{ij}' - \mu_i' \otimes \mu_j'|_{TV} \leq Cov_{ij}^{\Omega(1)}$.

Given this exercise, independent rounding produces a cut y such that $\mathbb{E} G(y) \geq (1 - poly(1/d, \lambda_2)) \tilde{\mathbb{E}} G(x)$, which is what we wanted to show.