15. Sum of Squares proof of KKL theorem (I)

Lecturer: Per Austrin

http://www.csc.kth.se/~lauria/sos14/

Instances with large integrality gap for the Arora-Rao-Vazirani SDP *relaxations of* Sparsest Cut *are approximable for constant degree sum of square relaxations. Such instances do not have small cuts and this can be proved in a sum of square proof of constant degree, using a degree 4 proof of the KKL theorem.*

Given a graph $G = (V, E)$ on N vertices, the Sparsest Cut problem asks to find a cut $(S, \bar{S})$ for $\emptyset \neq S \subset V$ such that $\frac{E(S,\bar{S})}{|S||\bar{S}|}$ is minimum. Arora et al. (2009)[1] present an $O(\sqrt{\log N})$-approximation algorithm for the sparsest cut. The algorithm is based on a SDP relaxation of the problem, augmented with additional inequalities that enforce a metric structure to the semidefinite solutions. It turns out that the additional inequalities have sum of square proofs of constant degree, so the whole algorithm is essentially the integer rounding of the solution of a Lasserre relaxation of constant rank.

[1] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009

Closely related is the Balanced Separator problem in which we ask for the smallest cut such that both parts have roughly the same size. An approximation algorithm for Sparsest Cut gives an approximation algorithm for Balanced Separator with just a constant factor loss in the approximation.

A natural question is whether this approximation factor is optimum or whether the analysis of the algorithm can be improved. Devanur et al. (2006)[2] describe a family of graphs for Balanced Separator such that the SDP relaxation plus triangle inequalities on these graphs has an integrality gap of $\Omega(\log \log N)$. This disproves the original conjecture of Arora et al. (2009) that the actual gap was constant.

[2] Nikhil R Devanur, Subhash A Khot, Rishi Saket, and Nisheeth K Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 537–546. ACM, 2006

As discussed before, to prove such integrality gap it is sufficient to find an instance G (or rather a family of instances) with N vertices such that

1. The SDP relaxation of Balanced Separator has value $O(\frac{1}{\log N})$;
2. the integer solution has size at least $\Omega\left(\frac{\log \log N}{\log N}\right)$.

It is a legitimate question to ask whether the gap of constant rank Lasserre relaxations is tighter. A first step in this direction is to understand how Lasserre relaxations behave on the instances from Devanur et al. (2006).[3]

[3] We call them "DKSV instances".

O'Donnel and Zhou (2013)[4] show that the lower bound argument for the balanced separators of the DKSV instances is formalizable in sum of square proofs of degree 4. As a consequence degree-4 sum of square relaxations for the DKSV instances do have $O(1)$ integrality gap.

[4] R. O'Donnell and Y. Zhou. Approximability and proof complexity. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 13)*, pages 1537–1556, 2013

Theorem 1 (O'Donnel, Zhou)**.** *The degree-4 sum of square relaxation for* Balanced Separator *on DKSV instances has value* $\Omega\left(\frac{\log \log N}{\log N}\right)$.

Concepts from the analysis of boolean functions

Building a graph without small balanced cuts is easy (e.g., just take an expander) but here we need one for which the corresponding SDP has small value. Fix $n > 0$ and set $N = 2^n$, the DKSV graph is based on the hypercube of dimension n. The hypercube indeed has cuts (along each dimension) with exactly a $\frac{2}{n} = \frac{2}{\log N}$ fraction of the edges. To modify this construction for their purpose, the approach of Devanur et al. (2006) is to consider a group of actions on the vertices of the graph and then identify all vertices in same orbit. **The actual construction is not the topic of the lecture.** The main point of this lecture is that the proof that the graph has no small balanced separators can be formalized as a low degree sum of square proof that deals with analytic properties of boolean functions. Indeed a separator in the hypercube can be interpreted as a function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$, and the value of the cut is essentially the number of coordinates that make the value of function flip, averaging among all inputs.

We need some notation. For $x, y \in \{-1,1\}^n$ the vector $z = x \oplus y$ is the one such that $z_i = x_i y_i$. Vector $e_i \in \{-1,1\}^n$ is the one with is -1 in the i-th coordinate and 1 everywhere else. Thus $x \oplus e_i$ is the same as x but with the sign of the i-th coordinate flipped.

$x \oplus y$

Influence

Definition 2. *Consider a function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$. The* influence *of a coordinate $i \in [n]$ on f is defined as*

$$\mathrm{Inf}_i[f] = \Pr_{x\in\{-1,1\}^n}[f(x) \neq f(x \oplus e_i)]. \tag{1}$$

The total influence *of f is*

$$\mathrm{Inf}[f] = \sum_i \mathrm{Inf}_i[f] \tag{2}$$

A fundamental theorem in the analysis of boolean function gives a lower bound on the influence in term of the *variance* of the function, where the variance of f is

Variance

$$\mathrm{Var}[f] = \mathbb{E}_{x\in\{-1,1\}^n}[f(x)^2] - (\mathbb{E}_{x\in\{-1,1\}^n}[f(x)])^2$$

and in our case, since f has $-1, 1$ values this corresponds to

$$\mathrm{Var}[f] = 1 - (\mathbb{E}_{x\in\{-1,1\}^n}[f(x)])^2$$

Theorem 3 (KKL theorem)**.** [5] *For every function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ there is a coordinate $i \in [n]$ such that*

$$\mathrm{Inf}_i[f] \geq \Omega\left(\frac{\log n}{n}\right) \mathrm{Var}[f] \tag{3}$$

[5] J. Kahn, G. Kalai, and Nati Linial. The influence of variables on Boolean functions. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*, pages 68–80. IEEE, 1988

There is a function (the so-called "Tribes" function) for which all coordinates have influence $\frac{\log n}{n}(1 \pm o(1))$, so the theorem is tight.

Another tool for the analysis of boolean function is the study of the relation between q-norms. In particular we are interested in the relation of the q-norm of a function with respect to its second norm.

q-norms

Definition 4. *We define the q-norm of* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ *as*

$$\|f\|_q = \mathbb{E}_x[|f(x)|^q]^{\frac{1}{q}}. \tag{4}$$

If the function has small degree (i.e., no nonzero Fourier coefficients of large degree) then the behaviour of the q-norms for $q > 2$ is not too different from the one of the 2-norm.

Hypercontractive inequality (I)

Theorem 5. *For every* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ *of degree* d *and any* $q \geq 2$ *it holds that*

$$\|f\|_q \leq (\sqrt{q-1})^d \|f\|_2. \tag{5}$$

Noise operator

Another way to interpret the Hypercontractive inequality is through the *Noise Operator*. Given $x \in \{-1,1\}^n$ we let y be a noisy version of x, such that each coordinate y_i is correlated (but not necessarily identical) to the coordinate x_i. For $\rho \in [-1,1]$ we say that y is ρ-correlated with x if y is sampled according to the following distribution.

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1+\rho}{2} \\ -x_i & \text{with probability } \frac{1-\rho}{2}. \end{cases} \tag{6}$$

Then for any function $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ we define the linear operator

Noise operator

$$T_\rho f(x) = \mathbb{E}_y[f(y)] \tag{7}$$

where y is ρ-correlated to x. We observe that

$$T_1 f(x) = f(x) \qquad T_{-1} f(x) = f(-x) \qquad T_0 f(x) = \mathbb{E}_{z \in \{-1,1\}^n}[f(z)].$$

The natural interpretation of the noise operator is the one of a smoothing function. Indeed the effect of this function is to reduce the weight of the high degree Fourier coefficients. The following claim shows that the weight of a Fourier coefficient decreases exponentially with its degree.

Claim 6. *Fix* $\rho \in [-1,1]$ *and let* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$*, then for every* $S \subseteq [n]$,

$$\widehat{T_\rho f}(S) = \rho^{|S|} \widehat{f}(S). \tag{8}$$

Proof. T_ρ is a linear operator and $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$, so it is sufficient to compute $T_\rho \chi_S$.

$$T_\rho \chi_S = \mathbb{E}_y[\chi_S(y)] \tag{9}$$

for y which is ρ-correlated with x. This is equivalent to $\mathbb{E}_z[\chi_S(z \oplus x)]$ where z is -1 with probability $\frac{1-\rho}{2}$ and 1 with probability $\frac{1+\rho}{2}$, and where $z \oplus x$ is the coordinate-wise product of z and x. By definition $\chi_S(z \oplus x) = \chi_S(x)\chi_S(z)$, so $T_\rho \chi_S(x)$ can be written as $\chi_S(x) \cdot \mathbb{E}_z[\chi_S(z)]$, which means that the Fourier basis is a basis of eigenfunctions for the operator T_ρ, and that

$$\widehat{T_\rho f}(S) = \mathbb{E}_z[\chi_S(z)]\widehat{f}(S).$$

A simple computation shows that $\mathbb{E}_z[\chi_S(z)] = \rho^{|S|}$ and concludes the proof of the claim. □

The noise operator T_ρ allows another interpretation of the Hypercontractive inequality.

Hypercontractive inequality (II)

Theorem 7. *For every $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ of and $q \geq 2$ it holds that*

$$\left\| T_{\frac{1}{\sqrt{q-1}}} f \right\|_q \leq \|f\|_2. \tag{10}$$

or equivalently

$$\mathbb{E}_x \left[\left| T_{\frac{1}{\sqrt{q-1}}} f(x) \right|^q \right] \leq \mathbb{E}_x[|f(x)|^2]^{q/2}. \tag{11}$$

The two versions are equivalent. It is a simple exercise to prove that the second version implies the first version if f is assumed to be homogeneous. Consider a homogeneous polynomial f of degree d. Linearity of T_ρ and Claim 6 imply that $T_\rho f(x) = \rho^d f(x)$, so

$$\|T_\rho f\|_q = |\rho|^d \|f\|_q.$$

By setting $\rho = \frac{1}{\sqrt{q-1}}$ and applying the second version of the Hypercontractive inequality we get that

$$\|f\|_2 \geq \left\| T_{\frac{1}{\sqrt{q-1}}} f \right\|_q = \left(\frac{1}{\sqrt{q-1}} \right)^d \|f\|_q. \tag{12}$$

Corollary 8. *For any $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ and any $\rho \in [0,1]$, it holds that*

$$\left\| T_{\sqrt{\rho}} f \right\|_2^2 \leq \|f\|_{1+\rho}^2. \tag{13}$$

Proof. To prove this claim we need to use Hölder inequality[6].

$$\left\| T_{\sqrt{\rho}} f \right\|_2^2 = \langle T_{\sqrt{\rho}} f, T_{\sqrt{\rho}} f \rangle = \langle f, T_\rho f \rangle \leq \\ \leq \|f\|_{1+\rho} \cdot \left\| T_\rho f \right\|_{1+\frac{1}{\rho}} \leq \|f\|_{1+\rho} \cdot \left\| T_{\sqrt{\rho}} f \right\|_2 \tag{14}$$

where the first inequality is an application of Hölder inequality, and the second is an application of the Hypercontractive inequality on function $T_{\sqrt{\rho}} f$ with noise operator $T_{\sqrt{\rho}}$ and $q = 1 + \frac{1}{\rho}$. □

[6] Hölder inequality claims that

$$\langle f, g \rangle \leq \|f\|_p \cdot \|g\|_q$$

for any pair p, q such that $\frac{1}{p} + \frac{1}{q} = 1$ and $p, q > 1$. It also extends to the case in which $p = 1$ and $q = \infty$.

The *noise stability* of a function is the correlation between the function and its noisy version. In particular it is interesting to understand which functions are stable with respect to noise.

Definition 9. *The* noise stability *of a function* $f : \{-1,1\}^n \longrightarrow \mathbb{R}$ *is*

$$\mathsf{S}_\rho[f] = \langle f, T_\rho f \rangle = \mathbb{E}_{x,y}[f(x)f(y)] \tag{15}$$

where the expectation is taken for x *picked uniformly at random and* y *that is* ρ*-correlated with* x.

In order to give a sum of square proof of the KKL theorem, we will estimate the small set expansion of a particular weighted graph related to the hypercube, the so called *noisy hypercube*. For any $x, y \in \{-1,1\}^n$ we use the notation

Symmetric difference

$$x\Delta y \stackrel{\text{def}}{=} \{i \mid x_i y_i = -1\}.$$

The vertex set of the graph is $\{-1,1\}^n$, and the edge $x \sim y$ has weight equal to the probability to sample y which is ρ-correlated with x. Namely

$$\left(\frac{1-\rho}{2}\right)^{|x\Delta y|} \cdot \left(\frac{1+\rho}{2}\right)^{n-|x\Delta y|}.$$

Let $S \subseteq \{-1,1\}^n$, then $\mathsf{S}_\rho[\mathbb{1}_S]$ is essentially the weighted sum of the edges that go from a vertex in S to a vertex in S.

Theorem 10 (Small set expansion for the Noisy Hypercube)**.** *Let* $\rho \in [0,1]$ *and let* $f : \{-1,1\}^n \longrightarrow \{-1,0,1\}$*, then*

$$\mathsf{S}_\rho[f] \leq \mathbb{E}[f^2]^{\frac{2}{1+\rho}}. \tag{16}$$

Proof. We know that $\mathsf{S}_\rho[f] \leq \|f\|^2_{1+\rho}$ because of Corollary 8. Notice that since $f(x)$ is either -1, 0 or 1, then $|f(x)|^{1+\rho} = |f(x)|^2$ and thus

$$\|f\|^2_{1+\rho} = \mathbb{E}_x[|f(x)|^{1+\rho}]^{2/(1+\rho)} = \mathbb{E}_x[|f(x)|^2]^{2/(1+\rho)}.$$

□

Hypercontractive inequality for $q = 4$

In this section we prove the second version Hypercontractive inequality, for the simplest non trivial case which is $q = 4$.

Claim 11. *It holds that*

Hypercontractive inequality for $q = 4$

$$\left\| T_{\frac{1}{\sqrt{3}}} f \right\|_4 \leq \|f\|_2. \tag{17}$$

or equivalently that

$$\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} f\right)^4\right] \leq \mathbb{E}[f^2]^2. \tag{18}$$

Proof. The proof is by induction on the number of variables n. If $n = 0$ the function is constant and thus the inequality is trivially true.

If $n > 0$ we identify a variable x and write $f = x \cdot g + h$ where g, h are both functions not dependent on variable x. We can do this because f can be written as a multilinear polynomial.

The rest of the proof is just a chain of equations and inequalities. All expected values are taken on a random input $x \in \{-1, 1\}^n$.

$$\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} f\right)^4\right] \leq \mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} xg + T_{\frac{1}{\sqrt{3}}} h\right)^4\right] \leq$$
$$\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} xg\right)^4 + 4\left(T_{\frac{1}{\sqrt{3}}} xg\right)^3\left(T_{\frac{1}{\sqrt{3}}} h\right) + 6\left(T_{\frac{1}{\sqrt{3}}} xg\right)^2\left(T_{\frac{1}{\sqrt{3}}} h\right)^2 + 4\left(T_{\frac{1}{\sqrt{3}}} xg\right)\left(T_{\frac{1}{\sqrt{3}}} h\right)^3 + \left(T_{\frac{1}{\sqrt{3}}} h\right)^4\right] \tag{19}$$

since g does not depend on x, we get

$$T_{\frac{1}{\sqrt{3}}} xg = \left(T_{\frac{1}{\sqrt{3}}} x\right)\left(T_{\frac{1}{\sqrt{3}}} g\right) = \frac{x}{\sqrt{3}}\left(T_{\frac{1}{\sqrt{3}}} g\right) \tag{20}$$

which allows to write

$$\mathbb{E}\left[\frac{x^4}{9}\left(T_{\frac{1}{\sqrt{3}}} g\right)^4 + \frac{4x^3}{3\sqrt{3}}\left(T_{\frac{1}{\sqrt{3}}} g\right)^3\left(T_{\frac{1}{\sqrt{3}}} h\right) + 2x^2\left(T_{\frac{1}{\sqrt{3}}} g\right)^2\left(T_{\frac{1}{\sqrt{3}}} h\right)^2 + \frac{4x}{\sqrt{3}}\left(T_{\frac{1}{\sqrt{3}}} g\right)\left(T_{\frac{1}{\sqrt{3}}} h\right)^3 + \left(T_{\frac{1}{\sqrt{3}}} h\right)^4\right] \tag{21}$$

which we can simplify observing that the expectation is taken under the uniform distribution of inputs, so x is independent from g and h. In particular the expected value of x^d is 1 is d is even and 0 otherwise. In the end we get

$$\mathbb{E}\left[\frac{1}{9}\left(T_{\frac{1}{\sqrt{3}}} g\right)^4 + 2\left(T_{\frac{1}{\sqrt{3}}} g\right)^2\left(T_{\frac{1}{\sqrt{3}}} h\right)^2 + \left(T_{\frac{1}{\sqrt{3}}} h\right)^4\right] =$$
$$\frac{1}{9}\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} g\right)^4\right] + 2\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} g\right)^2\left(T_{\frac{1}{\sqrt{3}}} h\right)^2\right] + \mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} h\right)^4\right] \tag{22}$$

We apply Cauchy-Schwarz inequality on the central summand. **This is an operation we cannot formalize directly in sum of square proofs.** In the end we upper bound the previous expression by

$$\frac{1}{9}\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} g\right)^4\right] + 2\sqrt{\mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} g\right)^4\right] \cdot \mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} h\right)^4\right]} + \mathbb{E}\left[\left(T_{\frac{1}{\sqrt{3}}} h\right)^4\right]$$

The last expression has four expected values. All of them can be upper bounded using the Hypercontractive inequality, by induction on the number of variables.

$$\frac{1}{9}\mathbb{E}[g^2]^2 + 2\sqrt{\mathbb{E}[g^2]^2\mathbb{E}[h^2]^2} + \mathbb{E}[h^2]^2 \leq$$
$$\leq \mathbb{E}[g^2]^2 + 2\mathbb{E}[g^2]\mathbb{E}[h^2] + \mathbb{E}[h^2]^2 =$$
$$\left(\mathbb{E}[g^2] + \mathbb{E}[h^2]\right)^2 = \mathbb{E}[f^2]^2. \tag{23}$$

□

References

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009.

[DKSV06] Nikhil R Devanur, Subhash A Khot, Rishi Saket, and Nisheeth K Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 537–546. ACM, 2006.

[KKL88] J. Kahn, G. Kalai, and Nati Linial. The influence of variables on Boolean functions. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*, pages 68–80. IEEE, 1988.

[OZ13] R. O'Donnell and Y. Zhou. Approximability and proof complexity. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 13)*, pages 1537–1556, 2013.