From integrality gaps to hardness

We have seen how we can transform computational hardness results into integrality gaps. In a surprising work Raghavendra [2008] showed a transformation in the other direction. Namely, he showed how to take every constant degree integrality gap for a constraint satisfaction problem, and obtain a *hardness of approximation* result for the same problem with (essentially) the same parameters. Alas, there is one major fly in this ointment: the result is based on Khot's *Unique Games Conjecture*, on whose veracity there is no consensus.

We will discuss the Unique Games Conjecture, and the evidence for and against it, later in this course, but regardless of whether the conjecture is true or not, the techniques and ideas behind Raghavendra's result are beautiful, and have already found additional applications. One promising sign is a result of Chan [2013] , who gave a hardness of approximation result based merely on $P \neq NP$ which matches the parameters (and is inspired by) the degree $\Omega(n)$ integrality gap for "nice subspace predicates" we saw before. However, at the moment we still don't know of a generic transformation along these lines.

The *Max Cut* problem we saw before is an example of a *Constraint Satisfaction Problem (CSP)*. In such a problem, the instance I is given a list of functions $f_1, \ldots, f_m \colon \Sigma^n \to \{0,1\}$, where Σ is some finite set, and the goal is to find the assignment $x^* \in \Sigma^n$ that maximizes the fraction of i's such that $f_i(x^*) = 1$. This fraction is known as the *value* of the instance I, and is denoted as $\mathrm{val}(I)$.[1] For every subset $\mathcal{C}$ of functions that that map finite sequences of elements in Σ to $\{0,1\}$, we define $CSP(\mathcal{C})$ to be the class of CSP's where all constraints are in $\mathcal{C}$. One particular case of interest is when $\mathcal{C}$ the set $\mathcal{C}(\mathcal{P})$ where $\mathcal{P}$ is a finite set of functions mapping $\{0,1\}^k$ to $\{0,1\}$, and a function $f \colon \Sigma^n \to \{0,1\}$ is in $\mathcal{C}(\mathcal{P})$ if f is obtained by applying some function $P \in \mathcal{P}$ to k of its input symbols. The corresponding class of CSP's is denoted as $CSP(\mathcal{P})$.

[1] We will not make much distinction between algorithms whose goal is to compute the value and algorithms whose goal is to find the actual maximizing assignment. In practice most algorithms that are aimed at the former also achieve the latter.

For $1 \geq c > s \geq 0$, a (c,s)-*approximation algorithm* for a class $CSP(\mathcal{C})$ is an algorithm that outputs 1 on every $I \in CSP(\mathcal{C})$ such that $\mathrm{val}(I) \geq c$ and outputs 0 on every I such that $val(I) \leq s$.[2] A (c,s) *basic integrality gap* for a class $CSP(\mathcal{C})$ is an instance $I = (f_1, \ldots, f_m) \in CSP(\mathcal{C})$ such that $\mathrm{val}(I) \leq c$ but there is a pseudo-distribution μ of degree the maximum of $\deg f_i$ over $\{0,1\}^n$ such that $\tilde{\mathbb{E}}_\mu \sum_{i=1}^m f_i = m$. Raghavendra [2008] proved the following remarkable result:

[2] The reason for calling these parameters c and s is that c stands for "completeness" and s stands for "soundness". These names make the most sense in the context of hardness reductions, as we'll see below.

1. Theorem (Hardness of approximation from integrality gaps). *Assuming Khot's Unique Games Conjecture and $P \neq NP$, for every*

$c > s$ and set $\mathcal{P}$ of functions mapping Σ^k to $\{0,1\}$. If there exists a (c,s) degree $2|\Sigma|k$ integrality gap for $CSP(\mathcal{P})$ then for every $\epsilon > 0$ there is no polynomial time $(c-\epsilon, s+\epsilon)$ approximation algorithm for $CSP(\mathcal{P})$.

> **Note:** The actual semidefinite program that Raghavendra considered, which he called *Basic SDP* is weaker (i.e., contains fewer constraints) than degree $2|\Sigma|k$ sos, but stronger than degree 2 sos, though is arguably "morally" closer to degree 2 than degree $2|\Sigma|k$. Indeed, Basic SDP *is* a degree two sos relaxation of the maximization problem phrased in a somewhat different "constraint vs variable" formulation. In particular for the *Max Cut* problem, the Basic SDP formulation and the standard degree 2 sos formulation we saw before are equivalent.
>
> While it will not matter for our discussion beow, for the sake of completeness we describe the Basic SDP formulation for a general CSP $I = (f_1, \ldots, f_m)$ over alphabet Σ. It is the degree 2 sos relaxation for the problem of maximizing a polynomial F over $\{0,1\}^{n|\Sigma|+m|\Sigma|^k}$ where we use the $0/1$ variables $\{x_{i,\sigma}\}_{i\in[n],\sigma\in\Sigma}$ and $\{y_{\ell,\vec{\sigma}}\}_{\ell\in[m],\vec{\sigma}\in\Sigma^k}$. Intuitively, $x_{i,\sigma} = 1$ iff the i^{th} variable of the original assignment is σ, and $y_{\ell,\vec{\sigma}} = 1$ iff the k variables involved in the ℓ^{th} constraint have the assignment $\vec{\sigma}$.
>
> It is not too hard to come up with a quadratic polynomial F in these variables such that the maximum of $F(x,y)$ over all $(x,y) \in \{0,1\}^{|\Sigma|n+m|\Sigma|^k}$ will equal the maximum fraction of satisfiable constraints to the original CSP by an assignment in Σ^n. We leave verifying this as an **exercise**.

Since *max cut* is a particular instance of constraint satisfaction problem with degree two constraints, combining this with Feige-Schechtman's result we saw in the last lecture, we get the following theorem as a corollary:

2. Theorem (Hardness of approximation for max cut). *Let $\alpha_{GW} \sim 0.878$ and $x_{GW} \sim 0.845$ be the constants computed in the previous lecture. Assuming Khot's Unique Games Conjecture and $P \neq NP$, then for every $\epsilon > 0$ there is no polynomial time $(x_{GW} - \epsilon, \alpha_{GW} x_{GW} + \epsilon)$ approximation algorithm for $CSP(\mathcal{P})$.*

Theorem 2 was actually proven by Khot et al. [2004] before Theorem 1 and served as an inspiration to Raghavendra [2008] 's result. We will only show the proof of Theorem 2, in fact, only sketch it at that while indicating how it can be further generalized.

The Unique Games Conjecture

The *Unique Games Conjecture* was proposed by Khot [2002] . It concerns the following constraint satisfaction problem:

3. Definition (Unique Games). For every $1 \geq c^* > s^* \geq 0$ and $\ell \in \mathbb{N}$, the $UG_{c^*,s^*}(\ell)$ is the problem of distinguishing whether an instance of $CSP(\mathcal{P}_\ell)$ has value at least c^* or value at most s^*, where $\mathcal{P}_\ell$ is the set of 2-ary predicates on alphabet $[\ell]$ defined as:

$$\mathcal{P}_\ell = \left\{ P\colon [\ell]^2 \to \{0,1\} \;\middle|\; \forall x \in [\ell] \exists \text{ unique } y \in [\ell] s.t. P(x,y) = 1 \right\}. \tag{1}$$

The conjecture is the following:

For every $\epsilon > 0$, there exists some ℓ such that for $UG_{1-\epsilon,\epsilon}(\ell)$ is NP hard.

The requirement of completeness less than 1 is inherent, as the following exercise shows:

4. Exercise. For every $\ell, s^* < 1$, give a polynomial-time algorithm for the $UG_{1,s^*}(\ell)$ problem.

Tight hardness of approximation for Max Cut

We will not show the full proof of Theorem 2 (let alone Theorem 1), but we will illustrate some of the ideas behind it. A priori it seems very strange that a result like that could be proved. A (c,s) integrality gap is some finite mathematical object with particular properties. How can the existence of such an object prevent the existence of an efficient algorithm?

The idea is that such an integrality gap can be used as a *gadget* in a reduction from the (conjectured to be) hard computational problem $UG_{1-\epsilon,\epsilon}(\ell)$ into an instance of $CSP(\mathcal{P})$. The Unique Games conjecture posits that for some particular values $c^* = 1-\epsilon$ and $s^* = \epsilon$ it is computationally hard (specifically NP-hard) to distinguish, given an instance I of $UG_{1-\epsilon,\epsilon}(\ell)$, between the case that $val(I) \geq c^*$ and the case that $\mathrm{val}(I) \leq s^*$. The reduction we are looking for is some efficient map φ mapping an instance I of $UG_{1-\epsilon,\epsilon}(\ell)$ into an instance $\varphi(I)$ of $CSP(\mathcal{P})$ satisfying:

- **Completeness:** If $\mathrm{val}(I) \geq c^*$ then $\mathrm{val}(\varphi(I)) \geq c - \epsilon$.
- **Soundness:** If $\mathrm{val}(I) \leq s^*$ then $\mathrm{val}(\varphi(I)) \leq s + \epsilon$

We will not give a full description of the reduction, but will only mention some of its key features. Recall that the alphabet of I in $UG_{1-\epsilon,\epsilon}(\ell)$ is the set $[\ell] = \{1,\ldots,\ell\}$. The reduction will use as a gadget an error correcting code ECC mapping $[\ell]$ to $\{0,1\}^L$ for some

L, and will map an instance I of Π that uses n variables into a max-cut instance I' on nL vertices that are divided into n L-sized blocks. If $x \in [\ell]^n$ is an assignment that achieves value at least c^* for the instance I, then we obtain a bipartition with cut value $c - \epsilon$ for I' by cutting the L variables of the i^{th} block according to the L-length string $ECC(x_i)$. The particular code we will use is known as the *long code*. It is the map $ECC\colon [\ell] \to \{0,1\}^L$ for $L = 2^\ell$, for every $i \in [\ell]$ and $w \in \{0,1\}^\ell$, we define $ECC(i)_w = w_i$. It will be more convenient for us to think of (potentially corrupted) codewords of ECC as functions f mapping $\{0,1\}^\ell$ to $\{0,1\}$, where the actual (i.e., non corrupted) codewords correspond to the *dictator* functions of the form $f_i(w) = w_i$.

The Max-Cut gadget desiderata

To show that our reduction is *sound*, we need to show that given any bipartition $f\colon \{0,1\}^{nL} \to \{0,1\}$ of the vertices of I' that cuts more than $s + \epsilon$ fraction of the edges, we can *decode* it into an assignment $x \in [\ell]^n$ that satisfies at least an s^* fraction of the original constraints of I. It turns out that the key property is to show a way how to decode the restriction of the bipartition f to any particular block into a particular symbol in Σ. Informally, the "gadget" we need for the reduction, is a graph H with vertex set $V = \{0,1\}^\ell$ having the following properties:

- **Completeness:** For every $i \in [\ell]$, the cut value of the bipartition corresponding to the dictator f_i is at least $c - \epsilon$.
- **Soundness:** For every $f\colon \{0,1\}^\ell \to \{0,1\}$ that is "far" from a dictator function, the value of the cut corresponding to the bipartition f is at most $s + \epsilon$.

Constructing a gadget from an integrality gap

Roughly speaking, the idea behind the construction is as follows. Recall that the $(x_{GW} - o(1), \alpha_{GW} x_{GW} + o(1))$ integrality gap was obtained by taking the graph G whose vertices are the vectors in the d dimensional unit sphere and we put an edge between two vertices $u, v \in \mathbb{R}^d$ if $\langle u, v\rangle \leq \rho_{GW} + \epsilon$ for $\rho_{GW} = 1 - 2x_{GW}$. Our gadget graph H over the Boolean cube $\{0,1\}^\ell$ will be inspired by G, in the sense that we connect $w, z \in \{0,1\}^\ell$ if their correlation (i.e., the fraction of coordinates they agree on minus the fraction of coordinates they

disagree on) is at most $\rho_{GW} + \epsilon$. Let us now try to analyze this graph on an intuitive level.

The *completeness* property is fairly straightforward. Indeed, if $f : \{0,1\}^\ell \to \{0,1\}$ is a "dictator" function of the form $f(w) = w_i$ then for a random edge (w, z), the probability that $w_i \neq z_i$ is at least $1/2 + (\rho_{GW} + \epsilon)/2 = x_{GW} + \epsilon/2$.

The *soundness* property is much subtler, not least because we did not even define what being "far from a dictator" means. Let us take one particular example. Suppose that $f\colon \{0,1\}^\ell \to \{0,1\}$ is a *linear threshold function* of the form $f(w) = 1$ if $\sum \alpha_i w_i > \tau$ and $f(w) = 0$ otherwise, for some coefficients $\alpha_1, \ldots, \alpha_\ell, \tau$. If all the α coefficients but one are zero then f is a dictator, and so one way of saying that f is "far" from being a dictator is that all the coefficients are of roughly equal magnitude (say the same up to some constant).[3] In this case we can use the central limit theorem to argue that $\sum \alpha_i w_i$ is roughly the same as a normal variable with the same mean and variance. So, the probability that for a random edge (w, z), the bipartition f will cut (w, z) in the sense $f(w) \neq f(z)$ is essentially the same as the probability that we can distinguish between two ρ-correlated normal variables, but by the same calculations that we've done before, this will be at most $\arccos(\rho)/\pi$ which in our case would be at most $\alpha_{GW} x_{GW}$.

[3] In Fourier analysis parlance, this is known as the property of f having *small maximum influence*.

More generally, to analyze the soundness of this gadget, Khot et al. [2004] used a powerful generalization of the central limit theorem known as the *invariance principle* (Mossel et al. [2005]).[4] Roughly speaking, the invariance principle means that if f is "far" from a dictator in some technical sense, then it cannot distinguish between the case that its input comes from the $0/1$ Bernoulli distribution or the Gaussian distribution with the same moments. But then if f would have too good of a cut value, that would refute the isoperimteric result that underlies the integrality gap. The invariance principle can be thought of as an *inverse theorem*, saying that only "nice" functions (i.e., dictators or functions close to them) can and do distinguish between the sphere (or Gaussian space) and the cube. This completes our (admittedly quite partial and sketchy) outline of the proof.

[4] In fact the papers were in the reverse order. The motivation behind (Mossel et al. [2005]) was precisely to complete the soundness analysis of (Khot et al. [2004]).

What's unique about unique games?

We have not seen the full reduction, and so can not at this point truly appreciate why it needs to rely on the unique games conjecture. Indeed, the *label cover* problem, which is superficially quite similar

to unique games, is in fact NP-hard (see Exercise 5 below). A priori one could perhaps hope that there is a "minor modification" of this reduction so it can handle the case where the original constraint satisfaction problem instance I is "non unique" and hence be based merely on $P \neq NP$.

On a technical level, the current proof relies on the uniqueness property to reduce the task of verifying that the original constraints were satisfied to the "code checking" task of verifying that any assignment that has good value for the gadget is related to a true codeword. One could hope to get a more sophisticated gadget that would enable such "consistency checking" as well. Indeed, the current best candidate approaches to proving the unique games conjecture boil down to coming up with such gadgets. There are some obstacles to such an approach, showing that it would require more significant modifications. First, there is a sub-exponential time algorithm for unique games (Arora et al. [2010]) which implies that any such reduction based on the *label cover* problem (which is believed to be exponentially hard) would need to use some kind of "powering" step with a polynomial blow up the instance size in addition to any gadget. It also shows that there is a sense in which the Unique Games problem is qualitatively easier than Label Cover.

Also, the relation between gadgets and integrality gaps is not yet fully understood. For example, while the current gadgets are based on degree two integrality gaps, it turns out that they are inherently *not* integrality gaps for higher degree, as the invariance principle itself (or, more accurately, close variants of it) has a constant degree sum of squares proof (Barak et al. [2012]). We do not know whether it means that a gadget whose soundness proof is based on the invariance principle cannot be used to obtain such NP hardness reductions.

We will return to the Unique Games Conjecture later in this course. Given current research, it seems that understanding its truth is closely coupled with the question of understanding the extent of the power of the sum of squares algorithm.

5. Exercise (NP hardness of non-unique games). One variant of the PCP Theorem is that for every $\epsilon > 0$ there is some k and some family $\mathcal{P}$ of predicates on $\{0,1\}^k$ such that it is NP hard to distinguish between $CSP(\mathcal{P})$ instances of value at most ϵ and $CSP(\mathcal{P})$ instances of value at least $1 - \epsilon$. Show that you can reduce the arity k of the constraints to 2, at the cost of increasing the alphabet. That is, show that for every k and such family $\mathcal{P}$ of predicates over $\{0,1\}^k$, there is some ℓ and a family $\mathcal{P}'$ of predicates over $[\ell]^2$, and an efficient

reduction R such that for every instance I of $CSP(\mathcal{P})$, $R(I)$ is an instance of $CSP(\mathcal{P}')$ satisfiying:

- **Completeness:** If $\mathrm{val}(I) \geq 1 - \epsilon$ then $\mathrm{val}(R(I)) \geq 1 - 10\epsilon$
- **Soundness:** If $\mathrm{val}(I) \leq \epsilon$ then $\mathrm{val}(R(I)) \leq 10\epsilon$.

This shows that if we drop the uniqueness constraints on the constraints on $[\ell]^2$ in the definition of $UG_{1-\epsilon,\epsilon}(\ell)$ then the unique games conjecture becomes a corollary of the PCP Theorem.[5]

[5] **Hint:** Given an instance I with n variables and m k-ary constraints, the reduction will create an instance I' with alphabet size $\ell = 2^k$ and $n + m$ variables $x_1, \ldots, x_n, y_1, \ldots, y_m$ where the variables $x_1, \ldots, x_n$ are as in the original constraint and the variable y_i encodes the assignment to the k variables participating in the i^{th} constraint.

References

Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *FOCS*, pages 563–572. IEEE Computer Society, 2010.

Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *STOC*, pages 307–326. ACM, 2012.

Siu On Chan. Approximation resistance from pairwise independent subgroups. In *STOC*, pages 447–456. ACM, 2013.

Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC*, pages 767–775. ACM, 2002.

Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? In *FOCS*, pages 146–154. IEEE Computer Society, 2004.

Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low in.uences invariance and optimality. In *FOCS*, pages 21–30. IEEE Computer Society, 2005.

Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *STOC*, pages 245–254. ACM, 2008.