Notes for Lecture 3

These notes are based on my survey paper [6]. L.T.

Some Consequences of the PCP Theorem

We have seen that the PCP Theorem is equivalent to the inapproximability of Max 3SAT and other constraint satisfaction problems. In this lecture we will see several reductions that prove inapproximability results for other problems.

1 Max 3SAT with Bounded Occurrences

We begin with a reduction from the Max E3SAT problem on general instances to the restriction of Max E3SAT to instances in which every variable occurrs only in a bounded number of clauses. The latter problem will be a useful starting point for other reductions.

For the reduction we will need *expander graphs* of the following type.

Definition 1 (Expander Graph) *Let $c > 0$ be a constant. An undirected graph $G = (V, E)$ is a c-expander if, for every subset $S \subseteq V$, $|S| \leq |V|/2$, the number of edges $e(S, V - S)$ having one endpoint in S and one in $V - S$ is at least $c|S|$.*

For our purposes, it will be acceptable for the expander graph to have multiple edges. It is easy to prove the existence of constant-degree 1-expanders using the probabilistic method. Polynomial-time constructible 1-expanders of constant degree can be derived from [1] or [5], and, with a smaller degree, from [3]. In a later class, we will prove the following result.

Theorem 2 (Explicit construction of expanders) *For every constant c there is a constant $d = d(c)$ and an algorithm that on, input an integer $n > d$, runs in time polynomial in n and output a regular graph of degree d with n vertices that is a c-expander.*

In the following, we use d to denote the constant $d(1)$ in Theorem 2.

Let now φ be an instance of 3SAT with n variables $x_1, \ldots, x_n$ and m clauses. For each variable x_i, let occ_i be the number of *occurrences* of x_i, that is, the number of clauses that involve the literal x_i or the literal $\bar{x}_i$. We write $x_i \in C_j$ if the variable x_i occurs in clause C_j. Notice that $\sum_{i=1}^{n} occ_i = 3m$. For each i, construct a 1-expander graph $G_i = (V_i, E_i)$ where V_i has occ_i vertices, one for each occurrence of x_i in φ. We denote the vertices of V_i as pairs $[i, j]$ such that x_i occurrs in C_j. Each of these graphs has constant degree d.

We define a new instance ψ of Max E3SAT with $N = 3m$ variables $Y = \{y_{i,j}\}_{i\in[n],x_i\in C_j}$, one for each occurrence of each variable in φ. For each clause of φ we put an equivalent clause in ψ. That is, if $C_j = (x_a \vee x_b \vee x_c)$ is a clause in φ, then $(y_{a,j} \vee y_{b,j} \vee y_{c,j})$ is a clause in ψ. We call these clauses the *primary clauses* of ψ. Note that each variable of ψ occurs only in one primary clause.

To complete the construction of ψ, for every variable x_i in φ, and for every edge $([i, j], [i, j'])$ in the graph G_i, we add the clauses $(y_{i,j} \vee \bar{y}_{i',j})$ and $(\bar{y}_{i,j} \vee y_{i',j})$ to ψ. We call these clauses the *consistency clauses* of ψ. Notice that if $y_{i,j} = y_{i',j}$ then both consistency clauses are satisfied, while if $y_{i,j} \neq y_{i',j}$ then one of the two consistency clauses is contradicted.

This completes the construction of ψ. By construction, every variable occurrs in at most $2d + 1$ clauses of ψ, and ψ has $M = m + 3dm$ clauses.

We now claim that the cost of an optimum solution in ψ is determined by the cost of an optimum solution in φ and, furthermore, that a good approximation algorithm applied to ψ returns a good approximation for φ. We prove the claim in two steps.

Claim 3 *If there is an assignment for φ that satisfies $m - k$ clauses, then there is an assignment for ψ that satisfies $\geq M - k$ clauses.*

PROOF: This part of the proof is simple: take the assignment for φ and then for every variable $y_{i,j}$ of ψ give to it the value that the assignment gives to x_i. This assignment satisfies all the consistency clauses and all but k of the remaining clauses. $\Box$

Claim 4 *If there is an assignment for ψ that leaves k clauses not satisfied, then there is an assignment for φ that leaves $\leq k$ clauses not satisfied.*

PROOF: This is the interesting part of the proof. Let $a_{i,j}$ be the value assigned to $y_{i,j}$. We first "round" the assignment so that all the consistency clauses are satisfied. This is done by defining an assignment b_i, where, for every i, the value b_i is taken to be the majority value of $a_{i,j}$ over all j such that $x_i \in C_j$, and we assign the value b_i to all the variables $y_{i,j}$. The assignment b_i satisfies all the consistency clauses, but it is possible that it contradicts some primary clauses that were satisfied by $a_{i,j}$. We claim that, overall, the b_i assignment satisfies at least as many clauses as the $a_{i,j}$ assignment. Indeed, for each i, if b_i differs from the $a_{i,j}$ for, say, t values of j, then there can be at most t primary clauses that were satisfied by $a_{i,j}$ but are contradicted by b_i. On the other hand, because of the consistency clauses being laid out as the edges of a 1-expander graph, at least t consistency clauses are contradicted by the $a_{i,j}$ assignment for that value of i alone, and so, the b_i assignment can be no worse.

We conclude that b_i assignment contradicts no more clauses of ψ than are contradicted by $a_{i,j}$, that is, no more than k clauses. When we apply b_i as an assignment for φ, we see that b_i contradicts at most k clauses of φ. $\Box$

In conclusion:

- If φ is satisfiable then ψ is satisfiable;

- If every assignment contradicts at least an ϵ fraction of the clauses of φ, then every assignment contradicts at least an $\epsilon/(1+3d)$ fraction of the clauses of ψ.

Theorem 5 *There are constants d and ϵ_2 and a polynomial time computable reduction from 3SAT to Max 3SAT-d such that if φ is satisfiable then $f(\varphi)$ is satisfiable, and if φ is not satisfiable then the optimum of $f(\varphi)$ is less than $1-\epsilon_2$ times the number of clauses. In particular, if there is an approximation algorithm for Max 3SAT-d with performance ratio better than $(1-\epsilon_2)$, then $\mathbf{P}=\mathbf{NP}$.*

2 Vertex Cover and Independent Set

In an undirected graph $G=(V,E)$ a vertex cover is a set $C \subseteq V$ such that for every edge $(u,v) \in E$ we have either $u \in C$ or $v \in C$, possibly both. An independent set is a set $S \subseteq V$ such that for every two vertices $u, v \in S$ we have $(u,v) \notin E$. It is easy to see that a set C is a vertex cover in G if and only if $V-C$ is an independent set. It then follows that the problem of finding a minimum size vertex cover is the same as the problem of finding a maximum size independent set. From the point of view of approximation, however, the two problems are not equivalent: the Vertex Cover problem has a 2-approximate algorithm (but, as we see below, it has no PTAS unless $\mathbf{P}=\mathbf{NP}$), while the Independent Set problem has no constant-factor approximation unless $\mathbf{P}=\mathbf{NP}$.

We give a reduction from Max E3SAT to Independent Set. The reduction will also prove intractability of Vertex Cover. If we start from an instance of Max E3SAT-d we will get a bounded degree graph, but the reduction works in any case. The reduction appeared in [4], and it is similar to the original proof of **NP**-completeness of Vertex Cover and Independent Set [2].

Starting from an instance φ of E3SAT with n variables and m clauses, we constuct a graph with $3m$ vertices; the graph has a vertex $v_{i,j}$ for every occurrence of a variable x_i in a clause C_j. For each clause C_j, the three vertices corresponding to the three literals in the clause are joined by edges, and form a triangle (we call such edges *clause edges*). Furthermore, if a variable x_i occurrs positively in a clause C_j and negated in a clause $C_{j'}$, then there is an edge between the vertices $v_{i,j}$ and $v_{i,j'}$ (we call such edges *consistency edges*). Let us call this graph G_φ. See Figure 1 for an example of this construction.

Note that if every variable occurrs in at most d clauses then the graph has degree at most $d+2$.

Claim 6 *There is an independent set of size $\geq t$ in G_φ if and only if there is an assignment that satisfies $\geq t$ clauses in φ.*

Proof: Suppose we have an assignment a_i that satisfies t clauses. For each clause C_j, let us pick a vertex $v_{i,j}$ that corresponds to a literal of C_j satisfied by a_i. We claim that the set of picked vertices is an independent set in G_φ. To prove the claim, we note that we picked at most one vertex from each triangle, so that we do not violate any clause edge, and we picked vertices consistent with the assignment, so that we could not violate any consistency edge.

For the other direction, suppose we have an independent set with t vertices. The vertices must come from t different triangles, corresponding to t different clauses. We claim that we can satisfy

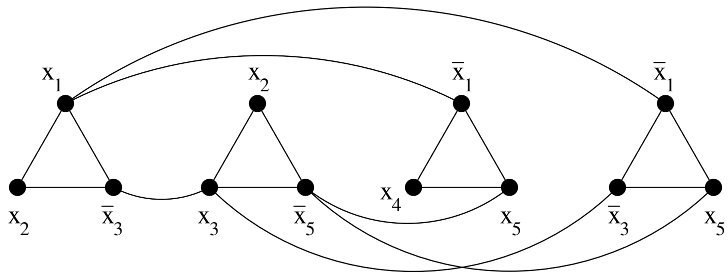

Figure 1: Graph construction corresponding to the 3CNF formula $\varphi = (x_1 \ \vee \ x_2 \ \vee \ \bar{x}_3) \wedge (x_2 \ \vee \ x_3 \ \vee \ \bar{x}_5) \wedge (\bar{x}_1 \ \vee \ x_4 \ \vee \ x_5) \wedge (\bar{x}_1 \ \vee \ \bar{x}_3 \ \vee \ x_5)$.

all such clauses. We do so by setting an assingment so that x_i takes a value consistent with the vertices $v_{i,j}$ in the independent set, if any. Since consistency edges cannot be violated, this is a well defined assignment, and it satisfies t clauses. $\Box$

If we combine this reduction with Theorem 5, we get the following result.

Theorem 7 *There is a polynomial time computable function mapping instances φ of 3SAT into graphs G_φ of maximum degree $d+2$ such that if φ is satisfiable then G_φ has an independent set of size at least $N/3$ (and a vertex over of size at most $2N/3$, where N is the number of vertices, and if φ is not satisfiable then every independent set in G_φ has size at most $N \cdot (1-\epsilon_2)/3$, and every vertex cover has size at least $N \cdot (2+\epsilon_2)/3$. In particular, if there is an approximation algorithm for Independent Set in degree-$(d+2)$ graphs with performance ratio better than $1/(1-\epsilon_2)$, or if there is an approximation algorithm for Vertex Cover in degree-$(d+2)$ graphs with performance ratio better than $1+\epsilon_2/2$, then $\mathbf{P} = \mathbf{NP}$.*

References

[1] O. Gabber and Z. Galil. Explicit construction of linear sized superconcentrators. *Journal of Computer and System Sciences*, 22:407–425, 1981. 1

[2] R.M. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. 3

[3] Alexander Lubotzky, R. Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988. 1

[4] C. H. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, 43:425–440, 1991. Preliminary version in *Proc. of STOC'88*. 3

[5] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):157–187, 2002. 1

[6] Luca Trevisan. Inapproximability of combinatorial optimization problems. Technical Report TR04-065, Electronic Colloquium on Computational Complexity, 2004. 1