High Dimensional Expanders

Lecture 12: Agreement Tests

Lecture by: Irit Dinur *Scribe: Irit Dinur*

In this lecture we describe agreement tests, which are a generalization of direct product tests and low degree tests, both of which play a role in PCP constructions. It turns out that the underlying structures that support agreement tests are invariably some kind of high dimensional expander, as we shall see.

1 General Setup

It is a basic fact of computation that any global computation can be broken down into a sequence of local steps. The PCP theorem [AS98, ALM+98] says that moreover, this can be done in a robust fashion, so that as long as *most* steps are correct, the entire computation checks out. At the heart of this is a *local-to-global* argument that allows deducing a global property from local pieces that fit together only approximately.

In an agreement test, a function is given by an ensemble of local restrictions. The agreement test checks that the restrictions agree when they overlap, and the main question is whether typical agreement of the local pieces implies that there exists a global function that agrees with most local restrictions.

Let us describe the basic framework, consisting of a quadruple $(V, X, \{\mathcal{F}_S\}_{S\in X}, \mathcal{D})$.

- **Ground set**: Let V be a set of n points (or vertices).
- **Collection of small subsets**: Let X be a collection of subsets $S \subset V$, typically for each $S \in X$ we have $|S| \ll n$.
- **Local functions**: for each subset $S \in X$, there is a space $\mathcal{F}_S \subset \{f : S \to \Sigma\}$ of functions on S. The input to the agreement test is an ensemble of functions

$$\{f_S \in \mathcal{F}_S \ : \ S \in X\}$$

- **Test distribution**: An agreement test is specified by giving a distribution $\mathcal{D}$ over pairs (or triples, etc.) of subsets S_1, S_2.

Given an ensemble $\{f_S\}$, the intention is that f_S is the restriction to S of a global function $F\colon V \to \Sigma$. Indeed, a local ensemble is called *perfect* if there is a global function $F\colon V \to \Sigma$ such that

$$\forall S \in X, \qquad f_S = F|_S.$$

An *agreement check* for a pair of subsets S_1, S_2 checks whether their local functions agree, denoted $f_{S_1} \sim f_{S_2}$. Formally,

$$f_{S_1} \sim f_{S_2} \qquad \iff \qquad \forall x \in S_1 \cap S_2, \quad f_{S_1}(x) = f_{S_2}(x).$$

A local ensemble which is perfect passes all agreement checks. The converse is also true: a local ensemble that passes *all* agreement checks must be perfect. We however will be interested in less-than-perfect ensembles, i.e. ensembles that pass a good fraction of the agreement checks, but perhaps not all of them.

An *agreement test* is specified by giving a distribution $\mathcal{D}$ over pairs (or triples, etc.) of subsets S_1, S_2. We define the agreement of a local ensemble to be the probability of agreement:

$$\text{agree}_{\mathcal{D}}(\{f_S\}) := \Pr_{S_1,S_2\sim D}[f_{S_1} \sim f_{S_2}].$$

An agreement theorem asserts that if $\{f_S\}_S$ is a local ensemble with $agree_{\mathcal{D}}(\{f_S\}) > 1-\varepsilon$ then it is close to being global.

2 Direct Product Tests

Perhaps the simplest agreement test to describe is the direct product test, in which X contains all possible k-element subsets of V. Namely,

- **Ground set**: $V = [n] = \{1, 2, \ldots, n\}$.
- **Collection of small subsets**: $X = \binom{[n]}{k}$ for some $k \ll n$. This notation stands for all possible k-element subsets of $[n]$.
- **Local functions**: For each $S \in X$, we let $\mathcal{F}_S$ be all possible functions on S, that is $\mathcal{F}_S = \Sigma^S = \{f : S \to \Sigma\}$
- **Test distribution**: There are several studied testing distributions. A central example is this: choose a parameter t such that $0 < t/k < 1$, and let $\mathcal{D}(t)$ be the distribution obtained by choosing a set T of t random elements in V, and then $S_1, S_2 \supset T$ uniformly and independently (such that $S_1, S_2 \in X$).

Suppose that $\text{agree}(\{f_S\}) \geq 1-\varepsilon$. Is there a global function $F : V \to \Sigma$ such that $F|_S = f_S$ for most subsets S? This is the content of the following theorem [DS14, DFH17]:

Theorem 2.1 (Direct Product Testing theorem)**.** *There exists constants $C > 1$ such that for all $\alpha, \beta \in (0,1)$ satisfying $\alpha + \beta \leq 1$, all positive integers $n \geq k \geq t$ satisfying $n \geq Ck$ and $t \geq \alpha k$ and $k - t \geq \beta k$, and all finite alphabets Σ, the following holds: Let $f = \{f_S : S \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying $agree_{\mathcal{D}}(f) \geq 1-\epsilon$, that is,*

$$\Pr_{(S_1,S_2)\sim\mathcal{D}(t)}[f_{S_1}|_{S_1\cap S_2} = f_{S_2}|_{S_1\cap S_2}] \geq 1-\epsilon.$$

Then there exists a global function $F : [n] \to \Sigma$ satisfying $\Pr_{S\in\binom{[n]}{k}}[f_S = F|_S] = 1 - O_{\alpha,\beta}(\epsilon)$.

The qualitatively strong aspect of this theorem is that in the conclusion, the global function agrees *perfectly* with $1 - O(\varepsilon)$ of the local functions. Achieving a weaker result where perfect agreement $f_S = F|_S$ is replaced by approximate one $f_S \approx F|_S$ would be significantly easier but also less useful. Quantitatively, this is manifested in that the fraction of local functions that end up disagreeing with the global function F is at most $O(\varepsilon)$ and is *independent of n and k*. It would be significantly easier to prove a weaker result where the closeness is $O(k\varepsilon)$ (via a union bound on the event that $F(i) = f_S(i)$).

3 Low Degree Tests

The first agreement test that was studied is the line vs. line [GLR+91, RS96] low degree test in the proof of the PCP theorem. A function $f : \mathbb{F}^m \to \mathbb{F}$ is said to have degree at most d if there is a polynomial

$$p(x_1, \ldots, x_m) = \sum_{(a_1,\ldots,a_m)\in\mathbb{F}^m:\ a_1+\ldots+a_m\leq d} \alpha_a \prod_{i=1}^{m} (x_i)^{a_i}$$

such that for all $z \in \mathbb{F}^m$, $p(z) = f(z)$. Low degree functions are useful in PCP constructions because of several reasons.

- Error correction: two low degree functions disagree on a large portion of their domain; in other words: the collection of low degree functions is an error correcting code (called the Reed-Muller code).
- Local testability: a low degree multivariate function remains low degree when restricted to a small-dimensional subspace. This gives a natural way to check locally whether a given function has low degree.
- Interpolation: Given an arbitrary function, it is easy to embed it into a low degree function through interpolation. This allows to encode any "proof" into a low degree function.

In the PCP construction, a low degree function on a large vector space is replaced by an ensemble of (supposed) restrictions to all possible affine lines. These supposed-restrictions are supplied by a prover and are not a priori guaranteed to agree with any single global function. This is taken care of by the "low degree test", which checks that supposed-restrictions on intersecting lines agree with each other, i.e. they give the same value to the point of intersection. The crux of the argument is the fact that the local agreement checks *imply* agreement with a single global function. Thus, the low degree test captures a local-to-global phenomenon.

- **Ground set**: $V = \mathbb{F}^m$ where $\mathbb{F}$ is a finite field.
- **Collection of small subsets**: X is the collection of all affine lines in V. An affine line is defined by fixing a pair of distinct points $x \neq y \in \mathbb{F}^m$, and then looking at the set

  $$S_{x,y} = \{t \cdot x + (1-t)y \ :\ t \in \mathbb{F}\}$$

  consisting of $|\mathbb{F}|$ points. We let X be the set of all such lines (observe that a line can be described through $\binom{|\mathbb{F}|}{2}$ different pairs of points $x \neq y \in S$).
- **Local functions**: For each $S \in X$, we let $\mathcal{F}_S$ be all low degree functions from S to $\mathbb{F}$ with degree at most d. Explicitly, for each $S = S_{x,y}$,

  $$\mathcal{F}_S = \left\{ f\colon S \to \mathbb{F} \;\middle|\; f(tx + (1-t)y) = \sum_{i=0}^{d} a_i t^i \text{ for some coefficients } a_0, \ldots, a_d \in \mathbb{F} \right\}.$$

  One can check that the definition of $\mathcal{F}_S$ does not depend on the choice of $x \neq y \in S$ because the set of low degree functions is invariant under affine transformations.
- **Test distribution**: Choose a random point $z \in \mathbb{F}^m$, and then independently choose two affine lines containing this point: $S_1, S_2 \ni z$.

The low degree testing theorem of [GLR+91, RS96] gives

Theorem 3.1. *For all finite fields $\mathbb{F}$ and all $d < |\mathbb{F}|/2$ and for all positive integers m, the following holds: Let $f = \{f_S \in \mathcal{F}_S \mid S \in X\}$ be an ensemble of local low degree functions satisfying $agree_{\mathcal{D}}(f) \geq 1 - \epsilon$, that is,*

$$\Pr_{(S_1,S_2)\sim\mathcal{D}}[f_{S_1}|_z = f_{S_2}|_z] \geq 1 - \epsilon,$$

where $z = S_1 \cap S_2$ is the point of intersection.
Then there exists a global function $F\colon V \to \mathbb{F}$ satisfying $\Pr_{S\in X}[f_S = F|_S] = 1 - O(\epsilon)$.

Furthermore, F itself is the evaluation of a polynomial function whose degree is at most d.

What is the role of "low degree"ness in this result? By requiring that the local functions have low degree, they are somewhat restricted. This makes the fact that two local functions agree on their intersection even more "surprising" and helps in proving the theorem. In contrast to the direct product setting, the theorem here simply fails if we change $\mathcal{F}_S$ to be Σ^S, as can be seen by the following example. For each S independently choose a random point $z_S \in S$ and let $f_S(z) = 0$ for all $z \neq z_S$ and $f(z_S) = 1$. (Check indeed that this function does not have low degree). It is not hard to check that agree$(\{f_S\}) > 1 - 2/|\mathbb{F}|$ yet there is no function $F : V \to \mathbb{F}$ that agrees with even a tiny fraction of the lines S. Indeed the best candidate would be the zero function $F \equiv 0$, but for each $S \in X$ we have disagreement: $\Pr_S[f_S = F|_S] = 0$.

3.1 Other variants

There are several variants to agreement tests in the low degree setting. Instead of lines one can consider planes, namely two-dimensional subspaces, as was done by Raz and Safra. In that case the test looks at two planes that intersect not in a point but in a line. Interestingly both distributions are very similar wrt agreement, see [BDN17]. Even higher dimensional subspaces are considered in other PCPs. One can also replace lines by curves of higher degree, and this appears in the classical proof of the PCP theorem.

4 Agreement tests and high dimensional expansion

Agreement tests on high dimensional expanders were studied in [DK17]. TBD

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, May 1998. (Preliminary version in *33rd FOCS*, 1992).

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998. (Preliminary version in *33rd FOCS*, 1992).

[BDN17] Amey Bhangale, Irit Dinur, and Inbal Livni Navon. Cube vs. cube low degree test. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, pages 40:1–40:31, 2017.

[DFH17] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Agreement tests on graphs and hypergraphs. *CoRR*, abs/1711.09426, 2017.

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *Proc. 58th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 974–985, 2017.

[DS14] I. Dinur and D. Steurer. Direct product testing. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 188–196, 6 2014.

[GLR+91] Peter Gemmell, Richard J. Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *STOC 1991*, pages 32–42, 1991.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, April 1996. (Preliminary version in *23rd STOC*, 1991 and *3rd SODA*, 1992).